

PRINCIPLES OF **MODERN RADAR**

ADVANCED TECHNIQUES



William L. Melvin, James A. Scheer (Editors)

Principles of Modern Radar

Principles of Modern Radar

Vol. II: Advanced Techniques

William L. Melvin

Georgia Institute of Technology

James A. Scheer

Georgia Institute of Technology



Edison, NJ
scitechpub.com



Published by SciTech Publishing, an imprint of the IET.
www.scitechpub.com
www.theiet.org

Copyright © 2013 by SciTech Publishing, Edison, NJ. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the web at copyright.com. Requests to the Publisher for permission should be addressed to The Institution of Engineering and Technology, Michael Faraday House, Six Hills Way, Stevenage, Herts, SG1 2AY, United Kingdom.

While the author and publisher believe that the information and guidance given in this work are correct, all parties must rely upon their own skill and judgement when making use of them. Neither the author nor publisher assumes any liability to anyone for any loss or damage caused by any error or omission in the work, whether such an error or omission is the result of negligence or any other cause. Any and all such liability is disclaimed.

Editor: Dudley R. Kay
Production Manager: Robert Lawless
Typesetting: MPS Limited
Cover Design: Brent Beckley

10 9 8 7 6 5 4 3 2 1

ISBN 978-1-891121-53-1 (hardback)
ISBN 978-1-61353-024-5 (PDF)

Dedicated to the many students of Georgia Tech's professional
education courses, who inspired this book's development;
and
to our families, for all of their support and understanding.

Brief Contents

Preface xv

Publisher Acknowledgments xviii

Editors and Contributors xx

1 Overview: Advanced Techniques in Modern Radar 1

PART I Waveforms and Spectrum

2 Advanced Pulse Compression Waveform Modulations and Techniques 19

3 Optimal and Adaptive MIMO Waveform Design 87

4 MIMO Radar 119

5 Radar Applications of Sparse Reconstruction and Compressed Sensing 147

PART II Synthetic Aperture Radar

6 Spotlight Synthetic Aperture Radar 211

7 Stripmap SAR 259

8 Interferometric SAR and Coherent Exploitation 337

PART III Array Processing and Interference Mitigation Techniques

9 Adaptive Digital Beamforming 401

10 Clutter Suppression Using Space-Time Adaptive Processing 453

11 Space-Time Coding for Active Antenna Systems 499

12 Electronic Protection 529

PART IV Post-Processing Considerations

13 Introduction to Radar Polarimetry 589

14 Automatic Target Recognition 631

15 Multitarget, Multisensor Tracking 669

PART V Emerging Techniques

16 Human Detection With Radar: Dismount Detection 705

17 Advanced Processing Methods for Passive Bistatic Radar Systems 739

Appendix A: Answers to Selected Problems 823

Index 829

Contents

Preface xv

Publisher Acknowledgments xviii

Editors and Contributors xx

1 Overview: Advanced Techniques in Modern Radar 1

- 1.1** Introduction 1
- 1.2** Radar Modes 2
- 1.3** Radar and System Topologies 5
- 1.4** Topics in Advanced Techniques 6
- 1.5** Comments 14
- 1.6** References 15

PART I **Waveforms and Spectrum**

2 Advanced Pulse Compression Waveform Modulations and Techniques 19

- 2.1** Introduction 19
- 2.2** Stretch Processing 26
- 2.3** Stepped Chirp Waveforms 40
- 2.4** Nonlinear Frequency Modulated Waveforms 48
- 2.5** Stepped Frequency Waveforms 58
- 2.6** Quadriphase Signals 70
- 2.7** Mismatched Filters 75
- 2.8** Further Reading 81
- 2.9** References 81
- 2.10** Problems 84

3 Optimal and Adaptive MIMO Waveform Design 87

- 3.1** Introduction 87
- 3.2** Optimum MIMO Waveform Design for the Additive Colored Noise Case 89
- 3.3** Optimum MIMO Design for Maximizing Signal-to-Clutter Ratio 95
- 3.4** Optimum MIMO Design for Target Identification 99
- 3.5** Constrained Optimum MIMO Radar 104

3.6	Adaptive MIMO Radar	109
3.7	Summary	113
3.8	Further Reading	114
3.9	References	114
3.10	Problems	115

4	MIMO Radar	119
4.1	Introduction	119
4.2	An Overview of MIMO Radar	121
4.3	The MIMO Virtual Array	122
4.4	MIMO Radar Signal Processing	124
4.5	Waveforms for MIMO Radar	135
4.6	Applications of MIMO Radar	138
4.7	Summary	142
4.8	Further Reading	143
4.9	References	143
4.10	Problems	145

5	Radar Applications of Sparse Reconstruction and Compressed Sensing	147
5.1	Introduction	147
5.2	CS Theory	150
5.3	SR Algorithms	166
5.4	Sample Radar Applications	183
5.5	Summary	196
5.6	Further Reading	196
5.7	Acknowledgments	197
5.8	References	197
5.9	Problems	207

PART II

Synthetic Aperture Radar

6	Spotlight Synthetic Aperture Radar	211
6.1	Introduction	211
6.2	Mathematical Background	214
6.3	Spotlight SAR Nomenclature	220
6.4	Sampling Requirements and Resolution	225
6.5	Image Reconstruction	234
6.6	Image Metrics	240

- 6.7 Phase Error Effects 244
- 6.8 Autofocus 250
- 6.9 Summary and Further Reading 253
- 6.10 References 255
- 6.11 Problems 257

7 Stripmap SAR 259

- 7.1 Introduction 259
- 7.2 Review of Radar Imaging Concepts 264
- 7.3 Doppler Beam Sharpening Extensions 271
- 7.4 Range-Doppler Algorithms 286
- 7.5 Range Migration Algorithm 305
- 7.6 Operational Considerations 318
- 7.7 Applications 327
- 7.8 Summary 330
- 7.9 Further Reading 331
- 7.10 References 332
- 7.11 Problems 333

8 Interferometric SAR and Coherent Exploitation 337

- 8.1 Introduction 337
- 8.2 Digital Terrain Models 342
- 8.3 Estimating Elevation Profiles Using Radar Echo Phase 344
- 8.4 InSAR Operational Considerations 359
- 8.5 InSAR Processing Steps 362
- 8.6 Error Sources 375
- 8.7 Some Notable InSAR Systems 382
- 8.8 Other Coherent Exploitation Techniques 386
- 8.9 Summary 392
- 8.10 Further Reading 392
- 8.11 References 393
- 8.12 Problems 397

PART III

Array Processing and Interference Mitigation Techniques

-
- 9 Adaptive Digital Beamforming 401
 - 9.1 Introduction 401
 - 9.2 Digital Beamforming Fundamentals 404

9.3	Adaptive Jammer Cancellation	419
9.4	Adaptive Beamformer Architectures	435
9.5	Wideband Cancellation	441
9.6	Summary	449
9.7	Further Reading	449
9.8	References	449
9.9	Problems	451
<hr/>		
10	Clutter Suppression Using Space-Time Adaptive Processing	453
10.1	Introduction	453
10.2	Space-Time Signal Representation	459
10.3	Space-Time Properties of Ground Clutter	472
10.4	Space-Time Processing	474
10.5	STAP Fundamentals	478
10.6	STAP Processing Architectures and Methods	483
10.7	Other Considerations	491
10.8	Further Reading	493
10.9	Summary	493
10.10	References	494
10.11	Problems	496
<hr/>		
11	Space-Time Coding for Active Antenna Systems	499
11.1	Introduction	499
11.2	Colored Space-Time Exploration	500
11.3	Interleaved Scanning (Slow-Time Space-Time Coding)	515
11.4	Code Selection and Grating Lobes Effects	517
11.5	Wideband MTI [12], [4]	520
11.6	Conclusion	524
11.7	Further Reading	525
11.8	References	525
11.9	Problems	526
<hr/>		
12	Electronic Protection	529
12.1	Introduction	529
12.2	Electronic Attack	533
12.3	EW-Related Formulas	545
12.4	EP Overview	553
12.5	Antenna-Based EP	554

- 12.6 Transmitter-Based EP 561
- 12.7 Exciter-Based EP 562
- 12.8 Receiver-Based EP 567
- 12.9 Signal Processor-Based EP 572
- 12.10 Data Processor-Based EP 576
- 12.11 Summary 581
- 12.12 Further Reading 584
- 12.13 References 584
- 12.14 Problems 585

PART IV Post-Processing Considerations

- 13 Introduction to Radar Polarimetry 589
 - 13.1 Introduction 589
 - 13.2 Polarization 594
 - 13.3 Scattering Matrix 601
 - 13.4 Radar Applications of Polarimetry 611
 - 13.5 Measurement of the Scattering Matrix 618
 - 13.6 Summary 622
 - 13.7 Further Reading 622
 - 13.8 References 623
 - 13.9 Problems 626
-

- 14 Automatic Target Recognition 631
 - 14.1 Introduction 631
 - 14.2 Unified Framework for ATR 633
 - 14.3 Metrics and Performance Prediction 634
 - 14.4 Synthetic Aperture Radar 638
 - 14.5 Inverse Synthetic Aperture Radar 652
 - 14.6 Passive Radar ATR 656
 - 14.7 High-Resolution Range Profiles 658
 - 14.8 Summary 661
 - 14.9 Further Reading 661
 - 14.10 References 662
 - 14.11 Problems 668
-

- 15 Multitarget, Multisensor Tracking 669
 - 15.1 Review of Tracking Concepts 669
 - 15.2 Multitarget Tracking 677

- 15.3** Multisensor Tracking 691
- 15.4** Summary 695
- 15.5** Further Reading 695
- 15.6** References 696
- 15.7** Problems 698

PART V **Emerging Techniques**

- 16** Human Detection With Radar: Dismount Detection 705
 - 16.1** Introduction 705
 - 16.2** Characterizing the Human Radar Return 710
 - 16.3** Spectrogram Analysis of Human Returns 719
 - 16.4** Technical Challenges in Human Detection 722
 - 16.5** Exploiting Knowledge for Detection and Classification 727
 - 16.6** Summary 729
 - 16.7** Further Reading 729
 - 16.8** References 730
 - 16.9** Problems 736
-

- 17** Advanced Processing Methods for Passive Bistatic Radar Systems 739
 - 17.1** Introduction 739
 - 17.2** Evaluation of the 2D-CCF for the Passive Radar Coherent Integration 747
 - 17.3** Direct Signal and Multipath/Clutter Cancellation Techniques 755
 - 17.4** Signal Processing Techniques for Reference Signal Cleaning and Reconstruction 766
 - 17.5** 2D-CCF Sidelobe Control 775
 - 17.6** Multichannel Processing for Detection Performance Improvement 791
 - 17.7** Summary 814
 - 17.8** Acknowledgments 815
 - 17.9** Further Reading 815
 - 17.10** References 815
 - 17.11** Problems 819

Appendix A: Answers to Selected Problems 823

Index 829

Preface

This is the second volume in the *Principles of Modern Radar* series. While the first volume, *Principles of Modern Radar: Basic Principles* provides fundamental discussions of radar operation, *Principles of Modern Radar: Advanced Techniques* discusses key aspects of radar signal processing, waveforms, and other important radar techniques critical to the performance of current and future radar systems. It will serve as an excellent reference for the practicing radar engineer or graduate student needing to advance their understanding of how radar is utilized, managed, and operated.

What this Book Addresses

Modern radar systems are remarkably sophisticated. They can be configured in numerous ways to accomplish a variety of missions. As a result, radar is a highly multidisciplinary field with experts specializing in phenomenology, antenna technology, receivers or transmitters, waveforms, digital design, detection, estimation and imaging algorithms, electronic protection, tracking, target identification, multi-sensor fusion, systems engineering, test and evaluation, and concepts of operation. In addition to tremendous advances in computing technology, a trend is afoot in radar to move the digitization step closer and closer to the antenna element. This places great emphasis on the importance of the collection approach, sensor topology, and the particular algorithms and techniques applied to the incoming data to produce a superior product.

Principles of Modern Radar: Advanced Techniques addresses this aforementioned trend and the most important aspects of modern radar systems, including quite current subtopics. Readers will find modern treatment of multi-input/multi-output (MIMO) radar, compressive sensing, passive bistatic radar, signal processing, and dismount/human detection via radar. The chapters are organized in five sections: waveforms and spectrum, synthetic aperture radar, array processing and interference mitigation techniques, post-processing considerations, and emerging techniques.

Why this Book was Written

We and radar practitioners are aware of many very fine single subject radar reference books that build from core principles with in-depth treatment, and most of them are referenced within this book for further reading. However, we and SciTech felt strongly that selected advanced radar topics could be gathered and organized logically into a single volume. Moreover, such a volume could incorporate textbook elements, most notably problem sets, for use within academic programs and training classes often taught, and necessarily so, within industry and government. Even practicing engineers engaged in self-study appreciate logical development of topics and problems with answers to test their understanding. Very few advanced radar books, however, are written in a textbook style and include problem sets. The chief impediment to the advanced radar textbook idea

has always been the unlikelihood of any one, two, or even three authors possessing such a broad, yet deep, knowledge of, and experience with, so many advanced radar subjects. We are very proud to say that the chapters in this volume are written by noted experts in the radar field, all of whom are active researchers in their areas of expertise and most of whom are also instructors of short courses for practicing engineers. We are thankful to each of the contributing authors who share our vision of a long-needed advanced radar book covering a diverse array of topics in a clear, coherent, and consistent framework. Their unwavering dedication to quality and content – evidenced by their multiple rewrites in response to reviews and the volume editors’ suggestions for improvements — inspires us all.

How the Content was Developed

Each chapter has also been thoroughly vetted for content and technical accuracy by outside radar experts who volunteered to take part in SciTech Publishing’s community review process. All of the chapters received multiple reviews at different phases in the development cycle, starting with chapter outlines and proceeding through multiple manuscript drafts. It is most evident that the quality of *Principles of Modern Radar: Advanced Techniques* has been tremendously improved by the selfless and enthusiastic work of the volunteer engineers, scientists, and mathematicians who invested their own time to review book chapters, sometimes individually and sometimes in related chapter sequences, all to help develop a high quality and long-lasting single source advanced radar book. The reviewers of the manuscript are gratefully acknowledged and listed by name in later pages of this opening section.

The History of the POMR Series

It should be no surprise that organizing and publishing a book of this nature is a significant and challenging undertaking. It is an interesting fact that the *Principles of Modern Radar* series evolved from the initial goal of a single book. From early reviews and the enthusiasm of chapter contributor candidates, the single book became two: *POMR: Basic Principles*, published in early 2010, and the planned “advanced applications and techniques”, which then became three. Why? The second volume had grown to over 30 planned chapters, and it quickly became apparent that we needed to divide the second volume into two distinct volumes: *Advanced Techniques* and *Radar Applications*. Over the past two years, as chapters were written, reviewed, and revised, *Advanced Techniques* edged slightly ahead in progress and became our primary focus over the past nine months. *Principles of Modern Radar: Radar Applications* therefore follows the issuance of this book.

Acknowledgements

As editors for this volume, we are very grateful to the SciTech Publishing team. We thank them for their support, professionalism, and certainly their patience. We are especially appreciative that the publisher, Dudley Kay, President and Editorial Director, set the highest expectations on book quality as his primary goal. Robert Lawless, Production Manager,

tracked, organized, and refined the many disparate elements to bring them together as a coherent and consistent whole. Brent Beckley, Sales and Marketing Director, helped gather and manage the unusually numerous volunteer reviewers as an explicitly stated “community effort” and consequently understood our content and audience objectives far in advance of publication.

Most importantly, we are thankful to our families for their patience, love, and support as we prepared materials, revised, reviewed, coordinated, and repeated. This book, in part, represents time away from the ones we love and would not have been possible without their understanding and willingness to support our passion for engineering.

To our Readers

We hope the reader will enjoy this book as much as we enjoyed putting it together. It should be clearly evident to all that read these pages that radar is an exciting, dynamic, and fruitful discipline. We expect the future of radar holds even more adventure and promise.

Please report errors and refinements. We know from the publication of the first volume, *POMR: Basic Principles*, that even the most diligently reviewed and edited book is bound to contain errors in the first printing. It can be frustrating to see such errors persist even in many subsequent printings. We have come to appreciate how committed and meticulous SciTech Publishing is about correcting errors, and even making subtle refinements, with each printing of the book. So, it remains a “community effort” to catch and correct errors and improve the book. You may send your suspected errors and suggestions to:

pomr2@scitechpub.com

This email will reach us and SciTech concurrently so we can confer and confirm the modifications gathered for scheduled reprints. You are always welcome to contact us individually as well.

Bill Melvin

Georgia Institute of Technology
Atlanta, GA
william.melvin@gtri.gatech.edu

Jim Scheer

Georgia Institute of Technology
Atlanta, GA
jim.scheer@gtri.gatech.edu

Publisher Acknowledgments

Technical Reviewers

SciTech Publishing gratefully acknowledges the contributions of the following technical reviewers, who selected chapters of interest and read each carefully and completely, often in multiple iterations and often with substantive suggestions of remarkable thought and depth. Taken in aggregate, the value of their reviews was beyond measure:

Mounir Adjrad – *University College London, UK*
Christopher Allen – *University of Kansas, USA*
Ron Aloysius – *Northrop Grumman Corporation, USA*
Chris Baker – *The Ohio State University, USA*
Greg Barrie – *Defence R&D Canada, Canada*
Lee Blanton – *Radar System Engineer, General Atomics Aeronautical, USA*
Shannon Blunt – *University of Kansas, USA*
Arik D. Brown – *Northrop Grumman, USA*
Daniel Brown – *Applied Research Laboratory, Penn State University, USA*
Ron Caves – *Senior Analyst, MDA Systems, Ltd., Canada*
Kernan Chaisson – *Captain, USAF (retired), USA*
Jean-Yves Chouinard – *Université Laval, Canada*
Carmine Clemente – *University of Strathclyde, UK*
Gregory Coxson – *Technology Service Corporation, USA*
G. Richard Curry – *Consultant, USA*
Antonio De Maio – *Università degli Studi di Napoli Federico II, Italy*
Patrick Dever – *Fellow Engineer, Northrop Grumman, USA*
John Erickson – *USAF, Wright-Patterson AFB, USA*
Gaspare Galati – *Tor Vergata University, Italy*
Martie Goulding – *Chief Radar Engineer–Airborne Systems,
MDA Systems, Ltd., Canada*
Fulvio Gini – *University of Pisa, Italy*
Tim Hagan – *Lead Measurement Engineer, JT3, LLC, USA*
Theodoris Kostis – *University of the Aegean, Greece*
Lorenzo Lo Monte, – *Radar Systems Engineer, University of Dayton
Research Institute, USA*
Khalil Maalouf – *Metron, Inc., USA*
Yasser M. Madany – *Alexandria University, Egypt*
Doug Moody – *Mercer University, USA*
Lee Moyer – *Chief Technology Officer, Technology Service Corporation, USA*
Brian Mulvaney – *Research Engineer, Georgia Tech Research Institute, USA*
Tony Ponsford – *Raytheon Canada Ltd., Canada*
Earl Sager – *Consultant, USA*
Alexander Singer – *Thales Canada, Canada*

Craig Stringham – *Brigham Young University, USA*
N. Serkan Tezel – *Istanbul Technical University, Turkey*
Kai-Bor Yu – *Boeing Company, USA*
David Zasada – *MITRE, USA*
Francesco Zirilli – *Professor, Sapienza Universita di Roma, Italy*

Editors and Contributors

Volume Editors



Dr. William Melvin

Volume editor-in-chief and multiple chapter author

William Melvin is Director of the Sensors and Electromagnetic Applications Laboratory at the Georgia Tech Research Institute and an Adjunct Professor in Georgia Tech's Electrical and Computer Engineering Department. His research interests include systems engineering, advanced signal processing and exploitation, and high-fidelity modeling and simulation. He has authored over 160 publications in his areas of expertise and holds three patents on adaptive radar technology. Among his distinctions, Dr. Melvin is a Fellow of the IEEE, with the follow citation: "For contributions to adaptive signal processing methods in radar systems." He received the Ph.D., M.S., and B.S. (with High Honors) degrees in Electrical Engineering from Lehigh University.



Mr. James A. Scheer

Associate volume editor and Chapter 1 – Overview: Advanced Techniques in Modern Radar

Jim Scheer has 40 years of hands-on experience in the design, development, and analysis of radar systems. He currently consults and works part time for GTRI and teaches radar-related short courses. He began his career with the General Electric Company (now Lockheed Martin Corporation), working on the F-111 attack radar system. In 1975 he moved to GTRI, where he worked on radar system applied research until his retirement in 2004. Mr. Scheer is an IEEE Life Fellow and holds a BSEE degree from Clarkson University and the MSEE degree from Syracuse University.

Chapter Contributors



Mr. David Aalfs

Chapter 9 – Adaptive Digital Beamforming

David Aalfs is a Principal Research Engineer and Head of the Advanced Techniques Branch within the Air and Missile Defense Division of GTRI. He has over 15 years of experience in digital beamforming and adaptive signal processing for sensor arrays. He is director of Georgia Tech's Adaptive Array Radar Processing short course and a feature lecturer in several other Georgia Tech short courses including "Principles of Modern Radar" and "Phased Array Antennas & Adaptive Techniques."



Mr. Mike Baden

Chapter 2 – Advanced Pulse Compression Waveform Modulations and Techniques

Mike Baden has 30 years of experience in radar system modeling and analysis. His principal areas of focus include radar waveforms, automatic target recognition, and battlefield obscurants. Recent research has included target modeling for ballistic missile defense, and multipath exploitation. He has authored or co-authored over 70 papers and reports, and lectures in several GTRI short courses.

**Dr. Kristin Bing***Chapter 14 – Automatic Target Recognition*

Kristin F. Bing is a GTRI research engineer with experience in signal and image processing applied to radar and medical ultrasound. Her current research efforts include space-time adaptive processing, ground moving target identification, and synthetic aperture radar. She regularly teaches in Georgia Tech short courses on various radar topics and is an active member of IEEE.

**Prof. François Le Chevalier***Chapter 11 – Space-time Coding for Active Antenna Systems*

François Le Chevalier is Scientific Director of Thales Air Operations Division, and Professor, “Radar Systems Engineering”, at Delft University of Technology (The Netherlands). Mr. Le Chevalier pioneered the French developments in adaptive digital beamforming and STAP radar systems demonstrations, and shared RF apertures concepts design and experimental validation. An author of many papers, tutorials, and patents in radar and electronic warfare, also active in most International Radar Conferences Program Committees, Prof. Le Chevalier is the author of a book on “Radar and Sonar Signal Processing Principles” published by Artech House in 2002, and editor of “Non-Standard Antennas”, published by Wiley in 2011.

**Dr. Fabiola Colone***Chapter 17 – Advanced Processing Methods for Passive Bistatic Radar Systems*

Fabiola Colone received the laurea degree in Communication Engineering and the Ph.D. degree in Remote Sensing from the University of Rome “La Sapienza”, Rome, Italy, in 2002 and 2006, respectively. Her research interests include passive coherent location (PCL), multi-channel adaptive signal processing, and space-time adaptive processing (STAP) with application to mono- and bi-static radar systems. She is involved in scientific research projects funded by the European Union, the Italian Space Agency, the Italian Ministry of Research, and the Italian Industry. Her research has been reported in over 60 publications in international technical journals and conference proceedings. Dr. Colone served in the technical committee of many international conferences on radar systems. She was in the organizing committee, as the Student Forum Co-Chair, of the IEEE 2008 Radar Conference (Rome, Italy). She is member of the Editorial Board of the International Journal of Electronics and Communications (AEÜ) (Elsevier) since October 2011.

**Mr. Daniel A. Cook***Chapter 6 – Spotlight Synthetic Aperture Radar*

Dan Cook has over 10 years of experience in the fields of synthetic aperture radar and sonar. He is a Senior Research Engineer at GTRI specializing in signal processing for coherent imaging and adaptive filtering with application to synthetic aperture imaging, interferometry, coherent change detection, and image quality assessment.

**Mr. Michael Davis***Chapter 4 – MIMO Radar*

Mike Davis has been a research engineer at GTRI’s Sensors and Electromagnetic Applications Laboratory since 2008. He was previously employed at the General Dynamics Michigan Research and Development center, which was formerly ERIM. His research interests involve the application of statistical signal processing techniques to problems in radar.



Dr. Lisa Ehrman

Chapter 14 – Automatic Target Recognition and Chapter 15 – Multitarget, Multisensor Tracking

Lisa Ehrman is the Chief Engineer in the Air and Missile Defense Division at GTRI. She received her Ph.D. in electrical engineering from Georgia Tech and has ten years of work experience, largely focusing on target tracking, sensor fusion, and systems engineering. In that time, Lisa has also co-authored twenty-five conference papers, three journal articles, and numerous technical reports.



Dr. Matthew Ferrara

Chapter 5 – Radar Applications of Sparse Reconstruction and Compressed Sensing

Matthew Ferrara earned the BS degree in Mathematics at the North Dakota State University (Fargo, North Dakota) and the Masters and Ph.D. degrees in Mathematics from Rensselaer Polytechnic Institute (Troy, New York). He is a research mathematician for Matrix Research, Inc. (Dayton, Ohio). His interests include inverse problems, computer vision, and radar signal processing. Under support from the Air Force Office of Scientific Research (AFOSR), he is currently engaged in research to address radar imaging problems which include unknown target/sensor motion, sparsely sampled data, and strongly scattering targets. He is a member of the Society for Industrial and Applied Mathematics (SIAM) and the Institute of Electrical and Electronics Engineers (IEEE).



Mr. Joseph R. Guerici

Chapter 3 – Optimal and Adaptive MIMO Waveform Design

J. R. Guerici has over 27 years of advanced technology development experience in industrial, academic, and government settings—the latter included a seven year term with Defense Advanced Research Projects Agency (DARPA) where he led major new radar developments. The author of over 100 technical papers and publications, including the book *Space-Time Adaptive Processing for Radar* (Artech House), he is a Fellow of the IEEE for “*Contributions to Advanced Radar Theory and its Embodiment in Real-World Systems*”, and is the recipient of the 2007 IEEE Warren D. White Award for “*Excellence in Radar Adaptive Processing and Waveform Diversity*”.



Dr. Sevgi Zübeyde Gürbüz

Chapter 16 – Human Detection With Radar: Dismount Detection

Sevgi Gürbüz is an Assistant Professor in the Electrical and Electronics Engineering Department at the TOBB Economics and Technology University, Ankara, Turkey and a senior research scientist at the Scientific and Technological Research Council of Turkey (TUBITAK) Space Technologies Research Institute, Ankara, Turkey. Previously, she was with the AFRL Sensors Directorate, Rome, NY as a Radar Signal Processing Research Engineer and holds degrees in Electrical Engineering from the Georgia Institute of Technology (PhD) and the Massachusetts Institute of Technology (BS, MEng). Her current research interests include human detection with radar, target detection, tracking and identification with radar networks, synthetic aperture radar, wireless sensor networks and satellite data processing.



Dr. Byron M. Keel

Chapter – Advanced Pulse Compression Waveform Modulations and Techniques

Byron Keel is a Principal Research Engineer and Head of the Signals and Systems Analysis Branch within the Radar Systems Division of GTRI. He received his BSEE, MSEE, and PhD from Clemson University. He has over 23 years of experience in radar waveform design, signal processing, and systems analysis. He regularly teaches in GTRI sponsored short courses including “Principles of Modern Radar” and is course director and principal lecturer in “Radar Waveforms.”



Dr. Pierfrancesco Lombardo

Chapter 17 – Advanced Processing Methods for Passive Bistatic Radar Systems

Pierfrancesco Lombardo received the laurea degree in electronic engineering and the Ph.D. degree from the University of Rome “La Sapienza,” Rome, Italy, in 1991 and 1995, respectively. His main research interests are in adaptive radar signal processing, radar clutter modelling, mono- and multi-channel coherent radar signal processing, SAR image processing and radio-location systems. In such areas, he has been Project Manager of a number of research project funded by the European Union, the Italian Space Agency, the Italian Ministry of Research and the Italian radar industry. His research has been reported in over 200 publications in international technical journals and conferences. Prof. Lombardo is member of the IEEE AES Radar System Panel, the Editorial board of IET Proceedings on Radar Sonar & Navigation, and is associate Editor for Radar Systems for IEEE Trans. on Aerospace and Electronic Systems since June 2001. He is member of the Scientific Committee of SESAR (Single European Sky ATM Research) European Commission & Eurocontrol.



Dr. Krishna Naishadham

Chapter 13 – Introduction to Radar Polarimetry

Krishna Naishadham received the M.S. degree from Syracuse University, and the Ph.D. from the University of Mississippi, both in Electrical Engineering, in 1982 and 1987, respectively. He served on the faculty of Electrical Engineering for 15 years at the University of Kentucky, Wright State University (tenured Professor), and Syracuse University (Adjunct Professor). In 2002, he joined Massachusetts Institute of Technology Lincoln Laboratory as a Research. In 2008, he joined Georgia Institute of Technology, where he is currently a Research Professor in the School of Electrical and Computer Engineering. His research interests include novel multifunctional antenna design, antenna miniaturization and electronic packaging for wireless handheld devices, wearable antennas and sensors, RFID integration of sensor nodes, and carbon nanotube based chemical sensors. Dr. Naishadham published four Book Chapters and over 150 papers in professional journals and conference proceedings on topics related to computational EM, high-frequency asymptotic methods, antenna design, EMC, materials characterization and wave-oriented signal processing. He is currently the Chair of the Joint IEEE AP/MTT Chapter at Atlanta and serves on the Technical Program Committee for the International Microwave Symposium. He served as an Associate Editor for the *Applied Computational Electromagnetics Society (ACES) Journal*, and is currently an Associate editor of the *International Journal of Microwave Science and Technology*.



Mr. Jason T. Parker

Chapter 5 – Radar Applications of Sparse Reconstruction and Compressed Sensing

Jason T. Parker received the B.S. and M.S. degrees in electrical and computer engineering from The Ohio State University, Columbus, in 2004 and 2006, respectively. Since 2006, he has been a research engineer with the Sensors Directorate of the U.S. Air Force Research Laboratory. He is concurrently pursuing the Ph.D. in electrical engineering at The Ohio State University. His research interests include compressive sensing, adaptive signal processing, and inverse problems, with applications to radar target detection and imaging. He is a member of IEEE and both the Eta Kappa Nu and Tau Beta Pi engineering honor societies.



Mr. Aram A. Partizian

Chapter 12 – Electronic Protection

Aram Partizian is a Senior Research Scientist at GTRI where he contributes to the design, development, and field-testing of advanced radar electronic warfare technologies. He has over 30 years of experience in radar and electronic warfare field, including software and system engineering roles at Raytheon Company prior to joining Georgia Tech. He earned a B.A. in Physics from Oberlin College in 1977.



Dr. Lee C. Potter

Chapter 5 – Radar Applications of Sparse Reconstruction and Compressed Sensing

Lee C. Potter received the B.E. degree from Vanderbilt University, Nashville, TN, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, all in electrical engineering. Since 1991, he has been with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus. Professor Potter is also an investigator with the OSU Davis Heart and Lung Research Institute. His research interests include statistical signal processing, inverse problems, detection, and estimation, with applications in radar and medical imaging. Prof. Potter is a two-time recipient of the OSU MacQuigg Award for Outstanding Teaching.



Dr. Mark A. Richards

Chapter 8 – Interferometric SAR and Coherent Exploitation

Mark Richards is a faculty member in Electrical and Computer Engineering at the Georgia Institute of Technology, teaching and conducting research in the areas of digital signal processing, radar signal processing, and high performance embedded computing. He was previously Chief of the Radar Systems Division in the Sensors and Electromagnetic Applications Laboratory of the Georgia Tech Research Institute (GTRI). He is lead editor of *Principles of Modern Radar: Basic Principles* (SciTech Publishing, 2010) and the author of *Fundamentals of Radar Signal Processing* (McGraw-Hill, 2005), as well as co-editor or contributor to four other books. He received his Ph.D. from Georgia Tech in 1982 and is a Fellow of the IEEE.



Dr. Teresa Selee

Chapter 14 – Automatic Target Recognition

Teresa Selee is a Research Scientist in the Adaptive Sensor Technology branch of GTRI's Systems and Electromagnetic Applications Laboratory (SEAL). Her areas of research include target tracking and discrimination, as well as adaptive radar signal processing algorithms. She gives lectures in the SMTI short course with the Georgia Tech Defense Technology Professional Education Program, and earned her PhD in Applied Mathematics from North Carolina State University.



Dr. Gregory A. Showman

Chapter 7 – Stripmap SAR

Greg Showman is a Senior Research Engineer and Head of the Adaptive Sensor Technology Branch in GTRI. He has over 25 years of experience in advanced RF sensor research and development, with an emphasis on the design and implementation of innovative signal processing techniques for radar imaging, multi-dimensional adaptive filtering, and electronic protection. He frequently teaches SAR in GTRI-sponsored short courses, including the "Principles of Modern Radar," and is co-director of "Fundamentals of SAR Signal Processing" and responsible for the stripmap lecture series.

Overview: Advanced Techniques in Modern Radar

William L. Melvin, James Scheer

Chapter Outline

1.1	Introduction	1
1.2	Radar Modes	2
1.3	Radar and System Topologies	5
1.4	Topics in Advanced Techniques	6
1.5	Comments	14
1.6	References	15

1.1 INTRODUCTION

Modern radar systems are highly complex, leveraging the latest advances in technology and relying on sophisticated algorithms and processing techniques to yield exceptional products. *Principals of Modern Radar* [1] is the first in a series, covering basic radar concepts, radar signal characteristics, radar subsystems, and basic radar signal processing. This text is the second in the series and contains advanced techniques, including the most recent developments in the radar community. Specifically, much of *Principles of Modern Radar: Advanced Techniques* discusses radar signal processing methods essential to the success of current and future radar systems. Applying these techniques may require specific hardware configurations or radar topologies, as discussed herein.

Principles of Modern Radar: Advanced Techniques focuses on five critical radar topics:

- **Waveforms and spectrum**, including advanced pulse compression techniques to provide high resolution or tailor the compressed waveform's impulse response; jointly optimized or adapted transmit waveforms with complementary receive processing; multi-input, multi-output (MIMO) radar leveraging advances in waveform generation and multichannel antenna technology; and, compressive sensing.
- **Synthetic aperture radar (SAR)** theory and processing techniques for stripmap, spotlight, and interferometric modes.
- **Array processing and interference mitigation techniques** based on multichannel processing methods, including adaptive digital beamforming (ADBDF) for interference suppression and space-time adaptive processing (STAP) for target detection in clutter,

as well as space-time coded apertures for mission-tailored beam patterns. Electronic protection considerations are also broadly discussed in this section.

- **Post-processing considerations**, including the application of polarimetry to enhance the radar product, automatic target recognition, and multitarget tracking.
- **Emerging techniques** for dismounted personnel target detection and passive radar processing strategies.

1.2 | RADAR MODES

Radar systems are designed to detect, locate, characterize, and, in some cases, track targets of interest. Radar applications and specific modes are diverse. For example, radars are used on aircraft, missiles, satellites, ships, ground vehicles, and tripods. They attempt to detect, locate, characterize, and possibly track aircraft, missiles, ships, satellites, personnel, metallic objects, moving ground vehicles, buried objects—even mold growing within building walls. With such a wide variety of radar platforms and targets, the process of taxonomizing specific radars and their goals is a daunting task. However, considering two primary radar *super* modes is often general enough to cover most radar objectives. The techniques in this text correspond to one or both of these modes:

- Moving target indication (MTI): the detection, location, characterization, and tracking of moving objects, such as missiles, aircraft, ground vehicles, and personnel (so-called dismounts).
- Imaging radar: the high-resolution estimation of the electromagnetic backscatter from stationary or moving objects that yields a spatial image of the target in one, two, or even higher dimensions. One-dimensional images are called high-range resolution (HRR) profiles, whereas two-dimensional views are called synthetic aperture radar (SAR) images. When the radar is stationary and the target is moving or when both platforms are moving, the corresponding imaging mode is usually called inverse synthetic aperture radar (ISAR).

In the MTI mode, *dots* on a display are the primary radar product. Figure 1-1 is an example of ground target detections on a topographical map obtained via a ground moving target indication (GMTI) airborne radar mode.

The quality of each dot is a result of the system design and signal processing applied to the received reflections from target and clutter as well as the system's ability to mitigate radio frequency interference (RFI). Radar detection is based on two models, or hypotheses: the null hypothesis, H_0 ; and the alternative hypothesis, H_1 . The null hypothesis presumes the target is not present in the chosen radar data, whereas the alternative hypothesis corresponds to the case of target signal embedded in interfering signals consistent with the null hypothesis (viz., clutter, jamming, other interference, and uncorrelated noise responses). Each of the signals under the null hypothesis case is stochastic: the complex envelope of the return is derived from a particular statistical distribution and follows a certain temporal behavior. For example, the return from a given clutter patch is commonly assumed to have a complex envelope drawn from a Rayleigh distribution (complex Gaussian voltage) and a voltage response that decorrelates over time according to a Billingsley model [2] for an overland collection or Gaussian correlation model over water [3]. Likewise, the target response is stochastic. The corresponding H_1 distribution typically appears



FIGURE 1-1 ■
GMTI radar
detections (called
dots) shown in local,
plan view co-
ordinates on
topological map as
typically seen from
an airborne
surveillance platform
(after [http://en
.wikipedia.org/wiki/
Joint_STARS](http://en.wikipedia.org/wiki/Joint_STARS)).

displaced relative to the null hypothesis condition due to a shift in the mean but is otherwise overlapping.

The overlap between the null and alternative hypothesis distributions leads to ambiguity in the decision-making process: a decision region (determined by a threshold setting) corresponding to one model may also lead to a false declaration of the opposite model. These false declarations are either false alarms (the alternative hypothesis is chosen when in fact no target is present) or missed detections (the null hypothesis is chosen when in fact a target is present). The optimal detector follows from the likelihood ratio test (LRT) and involves operations on collected radar signals (usually after some preprocessing); a sufficient statistic, $\psi(\mathbf{x})$, is a canonical detector formulation [4, 5]. Identifying the region where sufficient statistic outputs *likely* correspond to the alternative versus null hypotheses with a specified Type I error (false alarm rate) requires knowledge of the joint probability distributions under both hypotheses: $p_{\psi(\mathbf{x})|H_0}$ is the probability density function (PDF) for the null hypothesis, and $p_{\psi(\mathbf{x})|H_1}$ is the PDF for the alternative hypothesis. The decision region is typically chosen so that if $\psi(\mathbf{x}) > \eta$, where η is the detection threshold, the alternative hypothesis is chosen; otherwise, $\psi(\mathbf{x}) \leq \eta$, corresponds to selection of the null hypothesis.

Figure 1-2 depicts the detection process. The area under $p_{\psi(\mathbf{x})|H_1}$ to the right of η gives the probability of detection (P_D), whereas the area under $p_{\psi(\mathbf{x})|H_0}$ to the right of η gives the probability of false alarm (P_{FA}). As seen from this depiction, the two distributions overlap, and the only way to increase P_D is to lower η and accept a higher P_{FA} .

Alternately, one might ask if there is a strategy to increase the separation between the null and alternative hypothesis distributions. Generally, this increased separation can be achieved via the appropriate exploitation of the radar measurement space, or degrees of freedom (DoFs), and advanced processing methods like ADBF and STAP. The objective in exploiting DoFs is to identify a measurement space where the target and interference (e.g., clutter, jamming) are separable. For example, spatial and fast-time DoFs are used to efficiently mitigate the impact of wideband noise jamming on the detection of a target located in proximity to the jammer, but still at a slightly different angle of arrival. Advanced processing methods combine the measurement DoFs in the most effective manner possible

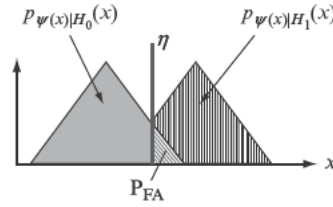
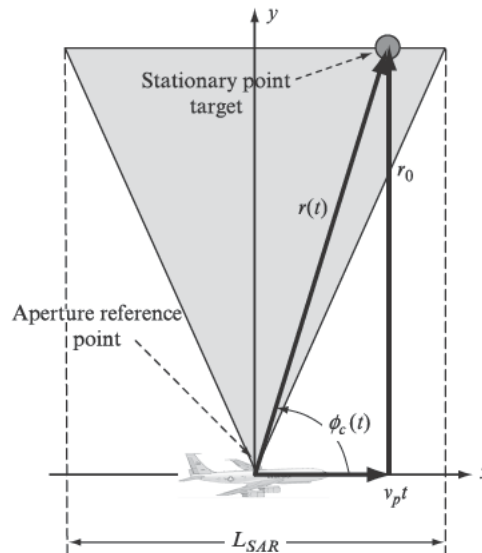


FIGURE 1-2 ■ Radar detection involves discriminating between the null (H_0) and alternative (H_1) hypotheses. This figure depicts H_0 and H_1 probability density functions for the sufficient decision statistic, along with threshold setting, η . The probability of false alarm, P_{FA} , is the area under the null hypothesis distribution curve to the right of the threshold, whereas the probability of detection is the area under the alternative hypothesis curve to the right of η .

to enhance MTI performance. The net objective of DoF selection and advanced processing methods in MTI radar is to increase the separation of the two distributions in Figure 1-2. Major sections of this text are devoted to examining these sophisticated techniques of critical importance to modern radar functionality.

The imaging radar mode typically involves moving the radar through angle while viewing a stationary target [6, 7]. (In the HRR case, a wideband waveform is used to characterize the target range response at that particular viewing angle.) As the radar moves through angle, the range between each of the various scatterers comprising the scene will vary in a manner consistent with the changing geometry. The changing range results in a time-varying phase that multiplies a complex gain term proportional to the square root of the scatterer's radar cross section (RCS). Each resolvable scattering cell in the unambiguous region of interest exhibits a unique phase history. Figure 1-3 depicts a SAR collection geometry, where L_{SAR} is the synthetic aperture length, r_o is the range from the aperture reference point to scene center, $r(t)$ is the time-varying range to a scatterer of interest, v_p is the platform velocity in the x-direction, t is the independent variable time, and $\phi_c(t)$ is the time-varying cone angle measured from the platform velocity vector aligned with the x-axis. From this figure, the reader can envision the time variation of $r(t)$ (or $\phi_c(t)$) as the platform moves along the synthetic aperture baseline.

FIGURE 1-3 ■ SAR collection geometry showing platform moving along x-axis with velocity, v_p , and a stationary point target passing through the gray illumination beam with time-varying range, $r(t)$. The platform is shown at the origin of the collection, and time varies from $-0.5T_{SAR}$ to $0.5T_{SAR}$, where T_{SAR} is the total collection time.



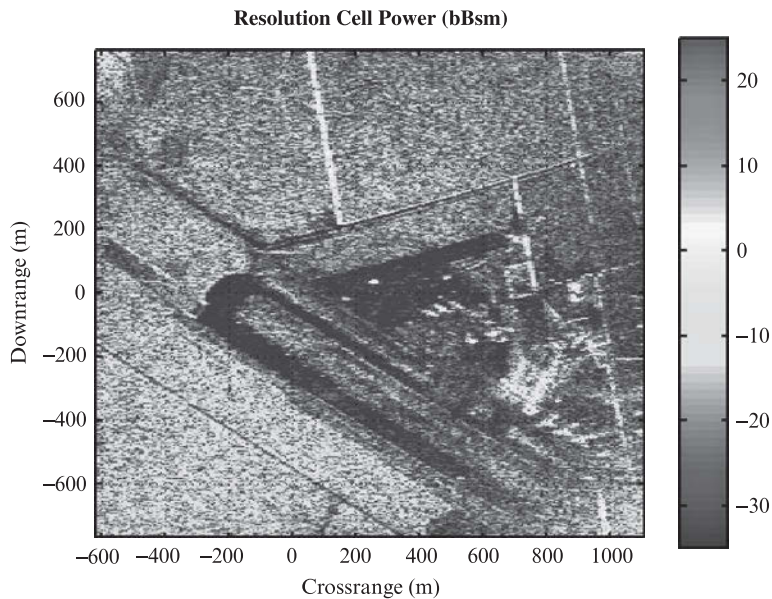


FIGURE 1-4 ■ Spotlight SAR image of Mojave Desert Airport at 1 m resolution, where bright areas indicate fence lines, sides of buildings, and aircraft on the tarmac (after [8], © 2004 IEEE).

The received radar signal is the summation of the returns from multiple, resolvable scatterers within the scene. (Unresolvable scatterers within each cell add coherently, yielding an effect known as *speckle* where some distributed scatterer responses appear brighter than others.) A matched filter designed to the phase history of a specified scattering cell, appropriately normalized and projected into the ground plane, yields an estimate of the corresponding RCS.

Figure 1-4 is an example of a 1 m spotlight SAR image collected at the Mojave Desert Airport in California, USA; the reader will notice features corresponding to tarmac, aircraft on the tarmac (bright spots on top of the darker regions), aircraft hangars, and fence lines. This image is plotted in the ground plane, where the x-axis corresponds to cross-range and the y-axis is downrange.

Precisely constructing the matched filter for each scatterer is reliant on perfect knowledge of the scene geometry, platform attitude, and hardware characteristics as well as correct assumptions on the scattering behavior (viz., no interaction between scattering cells consistent with the Born approximation). Errors in this knowledge lead to degraded image quality. Additionally, applying the precise matched filter can prove computationally burdensome. SAR algorithms focus on compensating for certain types of collection errors and approximating the matched filter to mitigate computational loading. Additional SAR goals can involve extracting additional information, such as the target height. The theory of imaging radar and important processing techniques and approaches to enhance image quality are discussed extensively in this text.

1.3 | RADAR AND SYSTEM TOPOLOGIES

Most fielded radar systems are monostatic: the transmitter and receiver are colocated, with the scattering phenomenology uniquely dependent on the angle of incidence and reflection being equal. In some cases, there may be the appearance of significant separation between

transmitter and receiver, yet the relative separation is small compared with the typical detection range; the phenomenology is still monostatic in nature. Over-the-horizon radar (OTHR) is an example of this case. Also, when the transmitter and receiver are located on different parts of an aircraft, this is considered monostatic.

In the bistatic radar topology [9], the transmitter and receiver are separated a considerable distance such that scattering phenomenology differs from the monostatic case. For aerospace bistatic systems, the ground clutter spectral characteristics also appear much more complicated than in the monostatic configuration. Bistatic radars also may be cooperative or noncooperative. A cooperative bistatic radar controls, manages, or selects its source of illumination. In contrast, a noncooperative bistatic radar, sometimes called a passive bistatic radar, employs transmit sources of opportunity, such as cell towers, television and radio transmitters, and other radar systems. While the bistatic radar may not control its source of illumination, modern radar technology still allows these systems to apply coherent signal processing methods.

Multistatic radar involves multiple receivers and possibly transmitters. Multistatic radar provides a diversity of spatial measurements, which can be used to minimize target fading, improve target geolocation [10], and possibly enhance target recognition. Because the multistatic radar can use multiple transmitters and receivers, it is sometimes considered a multi-input, multi-output (MIMO) configuration.

However, the typical MIMO configuration is usually monostatic in nature and involves transmitting different, ideally uncorrelated, waveforms from each antenna subaperture. The ability to coherently transmit different waveforms from each subaperture leads to spatial diversity on transmit, which effectively leads to a secondary phase modulation on the received target signal that can potentially improve target location performance. MIMO radar may also have some advantages for sparse arrays—dealing with timing and position uncertainty and possibly mitigating spatial ambiguity—and enhancing SAR coverage rates. Fully adaptive MIMO provides the opportunity for improved detection by attempting to match the illumination waveform to the target class of interest. MIMO is an area of current, active research within the radar community, and its benefits are still being benchmarked.

This text considers monostatic, bistatic, and MIMO radar configurations. Advances in processing technology and techniques are key enablers for bistatic and MIMO radar topologies and are also central to significant improvements in monostatic radar performance.

1.4 | TOPICS IN ADVANCED TECHNIQUES

This section provides brief commentary on the major contributions of this text.

1.4.1 Waveforms and Spectrum

Pulse compression waveforms are used in radar systems primarily to achieve the range resolution of a physically shorter pulse width while providing acceptable average power corresponding to the longer pulse. Low probability of intercept is another consideration. A number of modulations are available and are intended to provide the most appropriate ambiguity function for the application at hand. The ambiguity function characterizes the waveform range impulse response and its sensitivity to Doppler modulation. The waveform

resolution is inversely proportional to the waveform bandwidth. Achieving high resolution within receiver bandwidth and other hardware constraints is yet another driving factor.

Chapter 2, “Advanced Pulse Compression Waveform Modulations and Techniques,” describes in detail three classes of waveforms intended to provide high resolution while averting receiver bandwidth and/or analog-to-digital converter (ADC) limitations. These waveforms include stretch processing, stepped chirped, and stepped frequency. Stretch processing essentially starts the radar signal processing chain within the analog receive hardware, beating the incoming waveform with a modulation that converts range delay to spatial frequency. The digital processing stage applies an ADC operating at a lower sample rate, but fully covering the lower bandwidth spectrum corresponding to a particular range swath of interest, and a Fourier transform to pulse compress the data. In this case, swath width is traded for the higher resolution corresponding to the transmit bandwidth. Stepped chirp is a coherent waveform using a series of chirps of modest bandwidth and pulse width at offset transmit frequencies. Each chirp is transmitted at a chosen pulse repetition interval (PRI) and received by a radar front end matched to the chirp bandwidth and center frequency. The digital signal processor synthesizes a waveform generally corresponding to the concatenated bandwidth of all the received chirp signals. The stepped chirp approach thereby allows for very high resolution using radar hardware with much lower instantaneous bandwidth. Stepped chirp requires increased control over the radar oscillator and timing sequence and a modest increase in processing complexity. The range swath is limited by the chosen PRI, and target Doppler is another factor limiting performance. Stepped chirp has application to high resolution SAR systems.

Stepped frequency is also discussed in Chapter 2. The stepped frequency waveform is a modulation of choice in instrumentation radars. The waveform generator sends a series of narrowband frequencies through the transmitter for a specified target viewing angle. The narrowband receiver collects each frequency and reconstructs a waveform corresponding to the composite, much higher bandwidth signal. Stepped chirp waveforms are not especially Doppler tolerant, requiring compensation for any scatterer motion (e.g., turntable movement). Chapter 2 also covers waveforms of a particular bandwidth whose design or receive processing tailors the sidelobe response while minimizing signal-to-noise ratio (SNR) loss. This analysis includes nonlinear frequency modulated (NLFM) waveforms and mismatched filtering methods. Quadriphase coded waveforms are also examined as a means to manage spectral sidelobes and thus mitigate electromagnetic interference (EMI) among different electronic systems.

For decades, radar systems have applied adaptive signal processing within the receive signal processing chain. Constant false alarm rate (CFAR) algorithms are the prime example: they estimate the ambient disturbance power and then apply a threshold multiplier, which is a function of the CFAR method and number of training samples, to set a detection threshold that ideally leads to a design false alarm rate [11, 12]. ADBF and STAP are more recent examples, where the signal processor modifies spatial or spatio-temporal weights in response to changes in the interference or clutter environment in an attempt to maximize output signal-to-interference-plus-noise ratio (SINR). CFAR, ADBF, and STAP have improved radar performance immensely. Chapter 3, “Optimal and Adaptive MIMO Waveform Design,” considers extending the success of adapt-on-receive methods to the joint adaptation of both transmit and receive characteristics. As mentioned earlier, radar detection enhancement is largely dependent on choosing the appropriate radar DoFs and modifying the system response to the changing interference environment to instantaneously improve output SINR. Extending this idea to the transmit side suggests modifying

the waveform frequency, spatial, temporal, and polarimetric features. Chapter 3 discusses the approach to design jointly optimized transmit waveforms and receive processing to maximize SINR. The transmit waveform, for example, can be optimized to shape spectral content to avoid bands where interference is present or to place energy where a specific target response may be greatest. The adaptation of the transmit waveform can prove challenging, but in this era of readily available auxiliary data (e.g., map data, information on building layouts), knowledge-aided pseudo-optimizations may prove quite useful [13].

Chapter 3 generalizes the transmit waveform adaptation over the spatial domain through the appropriately configured vector formulation to handle MIMO configurations. The concept of MIMO radar from the system perspective is then discussed in further detail in Chapter 4, “MIMO Radar.” MIMO radar, as described herein, generally refers to a monostatic radar with the ability to transmit different waveforms from a number of antenna subapertures and collect all reflected transmissions with a multichannel receive array. Unlike Chapter 3, Chapter 4 focuses on deterministic waveforms with ideally low cross-correlation functions. Moreover, it explores the benefits of the additional phase diversity on transmit, which has the potential to enhance the system’s ability to resolve targets in angle. The benefits of these increased spatial DoFs have application to SAR and MTI radar: MIMO radar may, under the right circumstances, increase SAR area coverage rate and lead to potentially better minimum detectable velocity (MDV) for a fixed coverage rate in the MTI mode.

Chapter 5, “Radar Applications of Sparse Reconstruction and Compressed Sensing,” covers the last topic in the waveforms and spectrum section of this text. The idea behind compressed sensing theory is that a desired radar signal can be represented relatively sparsely—with a small number of basis functions—and that this compression can be achieved or enhanced through the measurement process. As presented in Chapter 5, the theory of compressed sensing presumes a linear signal model of the form $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, where \mathbf{y} is the vector of measurements, \mathbf{A} is a matrix whose columns represent the measurement bases, \mathbf{x} is the complex valued signal vector of interest, and \mathbf{e} is additive noise. For example, \mathbf{x} may be the vector of complex gain terms proportional to the square root of the reflectivity values of various points on the earth’s surface, the columns of \mathbf{A} then represent the unique phase history of each point, and \mathbf{y} is the vector of radar measurements to be converted into a radar image. Sparse reconstruction is focused on efficiently and accurately solving for the true value of \mathbf{x} through regularization. As emphasized in Chapter 5, sparse reconstruction is not compressed sensing; rather, compressed sensing combines sparse reconstruction with constraints on the measurement matrix. These constraints are often satisfied through randomization of the measured signal, for reasons described in mathematical detail within the chapter. The benefits of compressed sensing to modern radar include the potential to reduce the vast amount of data collected by the radar while still being able to generate a product comparable to that resulting from Nyquist sampled signals.

1.4.2 Synthetic Aperture Radar

SAR systems sample a particular, fixed scene and then employ signal processing methods to convert the measurements to estimates of the reflectivity of each resolvable pixel of interest. SAR can be applied to remote sensing (e.g., Earth resources management), military missions, and planetary exploration.

The two primary SAR modes are called stripmap and spotlight. The distinction is a result of the manner by which data are collected and processed; otherwise, the objective of

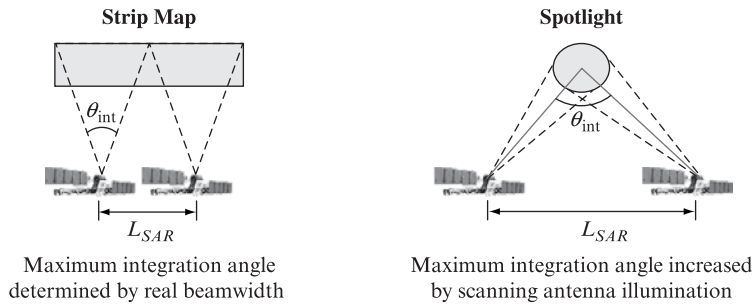


FIGURE 1-5 ■ Comparison of stripmap and spotlight SAR collection geometries, where L_{SAR} is the length of the synthetic aperture, and θ_{int} is the integration angle. In stripmap mode, the antenna beam “drags” through the scene of interest, whereas in spotlight mode the beam is continually re-steered to the center of the scene of interest.

each mode (viz., estimate the scene reflectivity) remains the same. Figure 1-5 shows the basic stripmap and spotlight mode collection geometries. The integration angle, the angle over which data are collected, is given as θ_{int} . SAR systems generally achieve down-range resolution consistent with the inverse of the transmit waveform bandwidth and cross-range resolution that is proportional to the ratio of the signal wavelength to twice the integration angle.

As Figure 1-5 indicates, the spotlight mode focuses a higher gain beam at a particular point on the earth’s surface. The beam is steered to the center of the scene as the platform takes samples over angle. The spotlight mode is the most popular when fine resolution is needed, since large integration angle is possible. Chapter 6, “Spotlight Synthetic Aperture Radar,” discusses spotlight imaging and corresponding algorithms. The primary viewpoint is that collected data represent the Fourier transform of the scene reflectivity. The polar formatting algorithm is a mainstay of spotlight image formation and is used to compensate for scatterer motion- through- resolution- cells (MTRC). Polar formatting resamples data collected along radial lines corresponding to each measurement angle onto a two-dimensional grid. Essentially, a two-dimensional inverse Fourier transform yields a radar image. Chapter 6 also explores multiplicative noise ratio (MNR), a key SAR metric that is a function of quantization noise, integrated sidelobe level, and ambiguity ratio. It varies as a function of signal strength in accordance with its constituent elements. Covered in this chapter also are the impact of phase errors and the most common autofocus methods used to improve image quality: phase difference autofocus and phase gradient autofocus. Autofocus is an adaptive method used to enhance image quality.

Stripmap mode and corresponding algorithms are discussed in Chapter 7, “Strip Map SAR.” The stripmap mode surveys the passing terrain using a sidelooking collection geometry. Stripmap mode has important application to large scene imaging for remote sensing (e.g., to examine deforestation, characteristics of polar ice, etc.). Chapter 7 discusses stripmap image formation algorithms in a sequence of increasingly sophisticated methods. The starting point is Doppler beam sharpening (DBS), which forms a range-Doppler map from the collected data over relatively small integration angle at long range and exploits the coupling between scatterer angle and Doppler frequency. Unfortunately, DBS image quality is limited by the occurrence of nonlinear phase as integration angle increases. Although the phase function is hyperbolic, an azimuth dechirp based nominally on a quadratic phase assumption is possible. Combining enhancements in range resolution with integration angle, range migration becomes a concern. DBS performance is extensible

to higher-resolution imaging by introducing range migration compensation and azimuth dechirp into the signal processing chain. However, higher-quality imagery requires better approximations to the individual point scatterer matched filter. Range-Doppler algorithms provide responses that more closely correspond to the scatterer point spread response (PSR), even precisely matching scatterer responses at certain ranges. Depth of focus—the range swath over which the PSR approximation yields acceptable image quality—is a primary limitation of such methods. The range migration algorithm (RMA) is presented as the culmination of the various stripmap SAR imaging formation methods discussed in this chapter. RMA makes no approximations to the PSR and is computationally efficient; it is the method of choice for higher-resolution stripmap imagery.

Interferometric SAR (InSAR or IFSAR) involves coherent exploitation of SAR imagery to derive terrain height information. Generally, terrain height is measured using pairs of SAR complex imagery (or multiple coherent collects) at slightly offset baselines, as described in Chapter 8, “Interferometric SAR and Coherent Exploitation.” The offset baseline provides diversity in range measurements as input into the InSAR terrain height estimation process. InSAR processing involves registration, phase unwrapping, and several other steps to calibrate the pixel height estimate. Airborne and spaceborne radar have successfully provided digital elevation maps (DEMs) for a number of years. Chapter 8 also describes other related techniques involving coherent exploitation of multiple, registered SAR collections, including coherent change detection and subsidence measurement.

1.4.3 Array Processing and Interference Mitigation Techniques

Section 1.2 suggests that measurement diversity and the ability to adapt to the changing characteristics of the interference environment are critical to enhanced detection and imaging performance.

Chapter 9, “Adaptive Digital Beamforming,” introduces the fundamentals of adaptive array radar technology. The concept of adapting an array of antennas to suppress interference dates to the late 1950s—with the work of Howells and Applebaum [14]—and has formed the basis for much of the field of adaptive signal processing. Advances in sensor and computing technology in recent years have led to increased emphasis on ADBF research and development.

Radar systems must provide adequate power-aperture to detect a target of a given RCS at a specified maximum range. Additionally, the radar must provide a mechanism to suppress interference and clutter. ADBF is used to suppress directional sources of RFI. The radar receive antenna design must include multiple spatial channels, which are used to discriminate the direction of arrival of a propagating electromagnetic wave. Digital beamforming uses the flexibility of digital signal processing to form multiple, simultaneous beams; the steering vector used to focus the array of antenna elements in a given direction corresponds to the spatial matched filter that maximizes output SNR. When colored noise is present, adapting the elements of the array to tailor the receive pattern is essential, as the RFI may be many orders of magnitude stronger than the target signal. ADBF attempts to maximize the array’s output SINR (the “I” indicates colored noise is present) by weighting array elements using estimates of the interference environment. The corresponding adaptive pattern, for example, will show nulls on sources of RFI—to within the limits of spatial channel availability—while forming a beam in the desired target direction. ADBF leads to significant performance improvement over the conventional solution when the target and RFI are sufficiently separated

in space. The required separation at which improvement is possible can be fractions of a beamwidth.

Chapter 9 comprehensively addresses ADBF theory and practical considerations. Multiple approaches to adapt the array are given, including the Wiener filter formulation; the maximum SINR weighting; constrained optimization, including the minimum variance distortionless response (MVDR) beamformer; the generalized sidelobe canceller, which is an approach to convert a constrained optimization into unconstrained form; and derivative and eigenvector constraints. Additionally, this chapter outlines a number of available approaches to calculate the weight vector in practice, including the batch sample matrix inverse (SMI) method and iterative methods. Element-level and subarray-based array architectures are explored, including key challenges associated with grating lobe effects. Chapter 9 also describes important hardware and computational considerations. The chapter culminates by describing several important adaptive antenna topologies, such as the sidelobe and beamspace cancellers, and considers methods for wideband cancellation based on space- and fast-time or sub-band architectures.

Chapter 10, “Clutter Cancellation Using Space-Time Adaptive Processing,” describes key issues in two-dimensional adaptive filtering using spatial and slow-time degrees of freedom to mitigate ground clutter. STAP is a generalization of ADBF techniques to two dimensions and is an important technology for aerospace radar searching for targets competing with stationary clutter reflections. This chapter formulates the space-time signal vector, discusses approaches to characterize space-time signals, and then develops a space-time ground clutter model. It is shown that ground clutter exhibits a distinct coupling in angle and Doppler; the STAP detection strategy is to identify signals whose angle-Doppler behavior differs from that of stationary clutter. In this vein, Chapter 10 then explains the essence of space-time processing, including key performance metrics such as probability of detection, SINR, SINR loss, and improvement factor. Several space-time adaptive algorithms are described as extensions of their one-dimensional counterparts given in Chapter 9. The chapter then covers STAP architectures, including reduced-dimension STAP and reduced-rank STAP. The reduced-dimension strategy is the most practical method of implementing STAP due to significant reduction in computational burden and training data requirements as well as performance benchmarking closely to the bound set by the joint-domain, optimal space-time processor. Benchmark results are given in the chapter using the SINR loss metric. A maximum likelihood estimator of target angle and Doppler response is given and is shown to integrate closely with the standard STAP solution. The chapter concludes with a summary of an end-to-end detection architecture and the practical issues of nonstationary or heterogeneous clutter impacts on STAP implementation.

It is important to point out that both ADBF and STAP exhibit super-resolution performance: they have the ability to null signals to within a fraction of a beamwidth, thereby providing acceptable performance even when the interference or competing clutter are within the mainlobe of the target beam. This makes them important in radar system design trades, where advanced signal processing coupled with modest size aperture replaces large, costly, conventional antenna systems.

Chapter 11, “Space-Time Coding for Active Antenna Systems,” describes space-time coding for multichannel arrays. This chapter primarily focuses on several limitations of the traditional approach to antenna transmit and receive—such as beamshape conflicts for multiple mission performance—and considers the flexibility afforded by transmitting different waveforms through separate spatial channels. In this sense, Chapter 11 combines

key MIMO elements of Chapters 3 and 4; however, it is unique in that it primarily focuses on array design issues in more detail than the prior chapters and provides several additional, practical applications. Moreover, this chapter looks at several waveform selections distinct from Chapter 4 and assesses their impact on array performance.

Chapter 12, “Electronic Protection,” discusses general strategies to protect the radar system from hostile denial or manipulation of the electromagnetic spectrum. It commences with detailed discussion of the two foremost classes of electronic attack (EA): noncoherent, or noise, jamming; and coherent jamming. A noncoherent jammer degrades radar sensitivity by injecting a noise-like waveform into the radar receiver; depending on the jammer configuration, the basic goal is to deny either the radar detection or range. A coherent jammer receives, delays, modulates, and retransmits the radar waveform; this EA approach takes advantage of radar signal processing gain, thus allowing the EA designer to employ much lower effective radiated power (ERP) than in the noncoherent jamming case. Coherent EA goals include masking and deception. A number of jammer deployments (e.g., stand-in, escort, distributed) are possible for both classes of jamming. Critical jammer formulae are subsequently given in Chapter 12. After the nature of EA is delineated, the goals and features of electronic protection (EP) are then comprehensively discussed. EP takes place throughout the radar implementation and can include the use of waveform diversity, low sidelobe or adaptive receive antenna designs, specialized signal processing methods, specific hardware designs, and variable radar concepts of operation (CONOPs). The EP attempts to deny the EA, including its electronic support (ES), key information needed to maximize effectiveness (e.g., operating frequency or waveform modulation) or to make the radar robust to the jamming signal (e.g., high dynamic range, adaptive jammer cancellation). The most effective EP methods are anticipatory, staying ahead of the deployed EA technique and thus minimizing degradation to the radar product in spectrally contested electromagnetic environments. Chapter 12 comprehensively discusses a variety of EP techniques. Adaptive digital beamforming, described in Chapter 9, is but one of many EPs; detailed understanding of ADBF from Chapter 9 is useful in comprehending aspects of the broad EP vantage of Chapter 12.

1.4.4 Post-Processing Considerations

Radar post-processing involves estimating target parameters, such as angle and range rate; placing the target in track; and determining the target class or specific features. Oftentimes, angle and Doppler estimation is considered part of the front-end radar signal processing, since it closely integrates with the antenna design and these parameter estimates may be used to mitigate false alarms (e.g., the processor may excise a detection if the corresponding angle-Doppler response is too close to the clutter ridge [13, 15]). In typical, modern radar design, the post-processor ingests target detections and parameters and tracks the target trajectory or accepts complex range profiles, SAR imagery, and Doppler spectra to automatically recognize the target type. This section is primarily focused on automatic target recognition (ATR) and multi-target tracking. Radar polarimetry is also discussed in this part of the text, as ATR serves as a primary driver of polarimetric diversity.

Chapter 13, “Introduction to Radar Polarimetry,” discusses the polarimetric properties of propagating waves and their interactions with radar targets. Key concepts include the various forms of linear, circular, and elliptical polarization and the polarization scattering matrix (PSM). The PSM is a complete description of the scattering object’s polarimetric

properties and is an important consideration in ATR, remote sensing, and system design to mitigate certain classes of clutter and interference.

A primer on target identification is given in Chapter 14, “Automatic Target Recognition.” Here a unified framework is given for ATR involving the following four steps:

- Identify the target set of interest.
- Select the feature set to enhance the probability of correct classification.
- Observe the feature set, which involves collecting the appropriate measurements to enhance target identification.
- Test the measurements for those features corresponding to a particular target or target class.

Example target features might include a specific engine modulation encoded onto the Doppler signature or a specific combination of target *bright spots* and polarimetric behavior in complex SAR imagery. Different target sets must exhibit different features in the measurement domain if the processor is to achieve acceptable target recognition performance. The radar employs a variety of strategies to collect measurements appropriate to separate the features of one type of target or class from another. HRR profiles, for example, measure the range response of the target and might uncover a specific distance between two dominant scatterers unique to that target class; fully polarimetric, complex SAR imagery encodes a number of details about the target that the processor correlates with library templates, where the shortest deterministic or statistical distance leads to a particular target declaration. And, as suggested earlier, the Doppler spectrum of an airborne target may disclose characteristics of a particular engine construction, hence revealing the target class. Chapter 14 considers each of these unified steps in extensive detail.

After a target has been detected, target parameter measurements—typically of target range, velocity, and angle—are assembled into tracks. The measurements are sometimes called *dots*, as they instantaneously appear as such on an operator display. The accuracy of each measurement is affected by the radar system design, target characteristics and geometry, and other environmental considerations such as clutter and interference residue. A challenging issue in target tracking is handling multiple, closely spaced targets. Chapter 15, “Multitarget, Multisensor Tracking,” discusses this important radar topic in detail. It introduces fundamental track concepts, including the interpretation of the track covariance and measurement-to-track association concepts. Track filtering involves propagating the state forward in time and then updating the state with a new measurement after the association step. The extended Kalman filter (EKF) is one track filtering method detailed in the early sections of the chapter. One of its limitations is that it applies the same propagation function to all targets, which may not be applicable to the multitarget environment. The multiple-hypothesis tracker (MHT) is used in multitarget tracking scenarios due to its ability to mitigate measurement-to-track association ambiguity; a significant portion of Chapter 15 is devoted to developing the MHT. Also, the interacting multiple model (IMM) is described as a way to mitigate mismatch between the presumed and actual target dynamical behavior. This chapter also covers multisensor tracking, which sometimes is also called track fusion.

1.4.5 Emerging Techniques

As a result of the maturation of subsystem technology—especially antenna and computing capability—the class of targets of interest in air-to-ground radar has quickly evolved

from large collections of vehicles to single large vehicles to personal conveyance to dismounts. Dismounts, as the name suggests, are walking or running humans. Chapter 16, “Human Detection with Radar: Dismount Detection,” explores methods to detect and characterize human targets. It first develops a time-varying, human RCS model. This model approximates the target response as the superposition of the returns from the head, torso, upper and lower arms, and upper and lower legs. The Thalman model characterizes target locomotion. The corresponding spectrogram of the dismount target is quite unique, exhibiting a time-varying sinusoidal Doppler response corresponding to the torso, with distinct, semiperiodic responses resulting from appendage reflections. The challenging aspect of the dismount response is that it is generally weak compared with typical ground vehicles. Moreover, the response time variation suggests that traditional approaches to pulse integration are not viable: as the energy smears over Doppler, a single Doppler hypothesis is inappropriate. On the positive side, though, the uniqueness of the dismount response is exploitable: the key is to employ model-based matched filters that search for plausible dismount returns in the collected radar measurements. Considering all possible dismount responses is a combinatorial challenge. Chapter 16 discusses practical matched filter strategies based on efficiently estimating dismount model parameters, which is extensible to dictionary-based approaches, such as orthogonal matching pursuit.

Passive bistatic radar (PBR), or passive coherent radar (PCR) as it is sometimes called, involves exploiting transmitters of opportunity—such as those on cell phone towers, carried by satellites, and used for direct broadcast communications—and, generally, lower-cost receivers to detect moving targets or image fixed scenes. The vast improvements in digital signal processing technology serve as the enabler for PCR. Chapter 17, “Advanced Processing Methods for Passive Bistatic Radar Systems,” discusses such PBR signal processing strategies. These primary steps include beamforming the reference channel and surveillance channel, mitigating clutter and interference, match filtering the surveillance channel using waveform information in the reference channel, and then forming and thresholding a range-Doppler map. System performance is determined by a number of factors, including the two-dimensional cross-correlation function (viz., the ambiguity function) for the passive waveform. This topic is considered at length, along with comprehensive discussion of practical PBR processing strategies and issues.

1.5 | COMMENTS

This text is generally organized by technical area, as described in Section 1.1 and summarized in Table 1-1, covering a number of contemporary topics. The topics primarily emphasize processing techniques that tend to serve as critical drivers in enhancing radar performance when combined with the appropriate measurement DoFs. Measurement DoFs set the physical limit on algorithm performance; the separation of target features, clutter response, and interference in the measurement domain is key to improved detection, estimation, and identification performance, thereby ultimately yielding better tracking capability. Electronic protection expands on the idea of exploiting measurement DoFs to all aspects of the radar design to provide resilience to electronic attack.

As seen from Table 1-1, this text broadly covers the most important, current, and emerging radar techniques. In this regard, *Principles of Modern Radar: Advanced Techniques* will serve as an invaluable reference for the student and radar practitioner.

TABLE 1-1 ■ Summary of Text Organization by Technical Area and Chapter

Technical Area	Chapters	Topics
Waveforms and spectrum	2, 3, 4, 5	Advanced pulse compression, MIMO techniques, compressive sensing
Synthetic aperture radar (SAR)	6, 7, 8	Stripmap SAR, spotlight SAR, interferometric SAR, imaging or coherent exploitation algorithms
Array processing and interference mitigation techniques	9, 10, 11, 12	Adaptive digital beamforming, space-time adaptive processing for clutter mitigation, space-time MIMO coded apertures for multimode radar, electronic protection
Post-processing considerations	13, 14, 15	Polarimetry, automatic target recognition, multitarget and multisensor tracking
Emerging techniques	16, 17	Human or dismount detection and characterization, passive bistatic radar processing methods

Each chapter ends with problem sets the interested reader may elect to solve. While a number of these may be solved in a traditional sense with pen and paper, many also require the use of MATLAB or another suitable programming language. With an emphasis on processing techniques, the best strategy to master the material herein is the hands-on approach.

1.6 | REFERENCES

- [1] Richards, M.A., Scheer, J.A., and Holm, W.A. (Eds.), *Principles of Modern Radar: Basic Principles*, SciTech Publishing, Raleigh, NC, 2010.
- [2] Billingsley, J.B., *Low-Angle Radar Land Clutter: Measurements and Empirical Models*, William Andrew Publishing, Inc., Norwich, NY, 2002.
- [3] Skolnik, M.I., *Introduction to Radar Systems*, 2d ed., McGraw Hill, New York, 1980.
- [4] DiFranco, J.V. and Rubin, W.L., *Radar Detection*, Artech House, Dedham, MA, 1980.
- [5] Johnson, D.H. and Dudgeon, D.E., *Array Signal Processing: Concepts and Techniques*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [6] Carara, W.G., Goodman, R.S., and Majewski, R.M., *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*, Artech House, Dedham, MA, 1995.
- [7] Sullivan, R.J., *Microwave Radar: Imaging and Advanced Concepts*, Artech House, Boston, MA, 2000.
- [8] Melvin, W.L., Showman, G.A., and Guerci, J.R., "A Knowledge-Aided GMTI Detection Architecture," *Proceedings of the IEEE Radar Conference*, April 26–29, 2004, Philadelphia, PA.
- [9] Willis, N. and Griffiths, H. (Eds.), *Advances in Bistatic Radar*, SciTech Publishing, Raleigh, NC, 2007.
- [10] Melvin, W.L., Hancock, R., Rangaswamy, M., and Parker, J., "Adaptive Distributed Radar," *Proceedings of the International Radar Conference*, October 2009, Bordeaux, France.

- [11] Nitzberg, R., *Radar Signal Processing and Adaptive Systems*, Artech House, Norwood, MA, 1999.
- [12] Levanon, N., *Radar Principles*, John Wiley & Sons, New York, 1988.
- [13] Melvin, W.L. and Guerri, J.R., “Knowledge-Aided Sensor Signal Processing: A New Paradigm for Radar and Other Sensors,” *IEEE Transactions on Aerospace and Electronic Systems*, July 2006, pp. 983–996.
- [14] Melvin, W.L., “Space-Time Adaptive Processing for Radar,” to appear in *Elsevier Electronic Reference in Signal, Image and Video Processing*, Academic Press.
- [15] Melvin, W.L. and Showman, G.A., “Knowledge-Aided, Physics-Based Signal Processing for Next-generation radar,” in *Proceedings of the Asilomar Conference on Signals, Systems, Computers*, November 2007, Pacific Grove, CA.

PART I

Waveforms and Spectrum

- CHAPTER 2** Advanced Pulse Compression Waveform Modulations and Techniques
- CHAPTER 3** Optimal and Adaptive MIMO Waveform Design
- CHAPTER 4** MIMO Radar
- CHAPTER 5** Radar Applications of Sparse Reconstruction and Compressed Sensing

Advanced Pulse Compression Waveform Modulations and Techniques

Byron Murray Keel, J. Mike Baden

Chapter Outline

2.1	Introduction	19
2.2	Stretch Processing	26
2.3	Stepped Chirp Waveforms	40
2.4	Nonlinear Frequency Modulated Waveforms	48
2.5	Stepped Frequency Waveforms	58
2.6	Quadrature Signals	70
2.7	Mismatched Filters	75
2.8	Further Reading	81
2.9	References	81
2.10	Problems	84

2.1 INTRODUCTION

This chapter surveys some of the more advanced pulse compression (PC) waveform modulations and techniques applied in modern radar systems, including stretch processing, stepped chirp waveforms, nonlinear frequency modulated (NLFM) waveforms, stepped frequency (SF) waveforms, quadrature codes, and mismatched filters (MMFs) applied to phase codes. Fundamentals of phase and frequency modulated PC waveforms are covered in [1].

In high range resolution systems, the waveform's instantaneous bandwidth places challenging requirements on the transmit/receive hardware and the signal processor. Stretch processing [2–6] is applied to a wideband linear frequency modulated (LFM) waveform to reduce the requirements levied on the analog-to-digital converter (ADC), data bus, and signal processor while maintaining the range resolution afforded by the transmit bandwidth.

Waveforms composed of narrowband pulses are used to accommodate bandwidth constraints imposed by hardware or processing limitations; however, in these instantaneously narrowband systems, fine range resolution may be achieved through interpulse frequency modulation. Waveforms exhibiting these properties include stepped chirp [7–15] and SF [16–18] waveforms.

LFM waveforms [1] exhibit unique properties that are advantageous in many systems and applications; however, a shortcoming is high range sidelobes. Amplitude tapers are applied to reduce the sidelobes at the expense of degraded resolution and a loss in a signal-to-noise ratio (SNR). NLFM waveforms [19–26] achieve low-range sidelobes without the need to apply an amplitude taper and without the corresponding loss in SNR.

The spectrum of a phase-coded waveform composed of rectangular-shaped chips exhibits significant energy outside the nominal waveform bandwidth [1]. The out-of-band energy is a source of electromagnetic interference (EMI). Quadriphase codes [27–31] are designed to reduce the spectral energy outside the passband while maintaining good range sidelobe performance and a near constant envelope.

A matched filter is commonly applied to a phase-coded waveform. The sidelobes of the range compressed response are thus governed by the code. However, in some cases, a mismatched filter [32–35] is used to further reduce or shape the response.

2.1.1 Organization

The chapter begins with an examination of stretch processing, which is used to reduce the sampling rate required in fine range resolution systems employing LFM waveforms. The technique is examined in Section 2 as well as performance metrics including data rate and throughput, range resolution, range window extent, processing gain, range-Doppler coupling, and sidelobes.

In Section 3, the stepped chirp waveform is explored as an alternative wideband LFM waveform with reduced transmit and receive intrapulse bandwidth requirements. The multiple pulse waveform and receiver architecture as well as techniques for stitching together the composite waveform within the signal processor are described. The waveform is shown to achieve a range resolution commensurate with the composite waveform bandwidth.

NLFM waveforms achieve low sidelobes via frequency modulation and avoid the SNR loss associated with applying an amplitude taper. Approaches for designing and synthesizing NLFM waveforms are examined in Section 4. In addition, examples are used to demonstrate achieved sidelobe performance and Doppler intolerance.

Stepped frequency waveforms are presented in Section 5 as signals composed of narrowband pulses, each with a different transmit center frequency, that are combined coherently to achieve fine range resolution. Processing steps, achieved range resolution, and waveform trades are explored.

Advanced phase code modulations and techniques are also examined. Quadriphase encoding is presented in Section 6 as a transformation applied to a biphasic code to reduce the waveform's spectral sidebands while preserving its range sidelobe performance. The lower sidebands reduce EMI. Mismatched filters are applied in Section 7 to reduce or tailor the sidelobe response of a phase-coded waveform. Sidelobe performance, processing loss, and Doppler tolerance are investigated.

2.1.2 Key Points

Important concepts developed throughout the chapter are summarized as follows:

- Stretch processing is a technique applied to a wideband LFM waveform to reduce the required processing bandwidth while maintaining the range resolution afforded by the transmit bandwidth.

- Stretch processing is applied in many high-resolution systems including synthetic aperture radars (SARs).
- A stepped chirp waveform consists of several LFM pulses that are shifted in frequency by a constant offset. The waveform provides a reduction in instantaneous bandwidth on both transmit and receive.
- The stepped chirp waveform is sampled at the single pulse bandwidth and is reconstructed in the signal processor to achieve a range resolution commensurate with the composite waveform bandwidth.
- NLFM waveforms employ frequency modulation to shape the spectrum and thereby reduce the range sidelobes. These waveforms do not require an amplitude taper as commonly used with an LFM waveform and thus avoid the associated SNR loss.
- NLFM waveforms are less Doppler tolerant than their LFM counterparts.
- Stepped frequency waveforms achieve a low instantaneous bandwidth on both transmit and receive. These narrowband signals are used to lessen hardware and processor requirements. Fine range resolution is achieved by stepping the frequency of each pulse to create a synthetic wideband waveform.
- An SF waveform is transmitted over a relatively long time interval, thus limiting its application; however, SF waveforms are used in some wideband instrumentation radar systems.
- Quadriphase codes exhibit low spectral sidelobes and thus help to mitigate EMI.
- Quadriphase codes are created by transforming a biphasic code into a four-phase code and applying a subpulse with a half-cosine shape. The subpulses overlap creating a nearly constant envelope.
- Mismatched filters may be designed and applied to phase codes to reduce the integrated and peak sidelobe responses with minimal loss in SNR. Mismatched filters may also be used to intentionally shape the sidelobe response.
- The performance of mismatched filters is sensitive to uncompensated Doppler shifts.

2.1.3 Notation

2.1.3.1 Common Variables

Variables used throughout the chapter are as follows:

t	time
j	$\sqrt{-1}$
π	3.14159265...
f_0	transmit center frequency
Ω	frequency in radians per second
c	speed of light
τ	pulse length
t_d	time delay associated with a point target
F_s	analog-to-digital converter sampling rate
ω	frequency in radians per sample
T_s	analog-to-digital converter sampling period
f_d	Doppler shift
Δf	frequency step size

T	pulse repetition interval
t_{gd}	group delay
δR	range resolution
ΔR	range window extent

2.1.3.2 Stretch Processing

Variables associated with stretch processing are as follows:

$x(t)$	transmit waveform
f_1	center frequency of first oscillator
β	LFM waveform's swept bandwidth
θ_1	phase of the first local oscillator
f_2	center frequency of the second oscillator
θ_2	phase of the second local oscillator
$LO_{2,tx}$	second oscillator signal used to synthesize transmit signal
t_{rcv}	time delay on receive, referenced to the center of the range window
$LO_{1,rcv}$	first oscillator signal applied on receive
$LO_{2,rcv}$	second oscillator signal applied on receive
$x_r(t)$	received signal
$y(t)$	received signal after mixer stages
Δt_d	time delay relative to the center of the range window
φ	residual video phase
θ	composite phase after deramp operation
f_b	beat frequency
$Y(\Omega)$	spectrum of the deramped signal
$Y_M(\Omega)$	spectrum magnitude
d	$d = -\tau(1 - \Delta t_d /\tau)$
Ω_{peak}	location of sinc's main lobe peak
Ω_{null}	location of sinc's first null
$\delta\Omega$	difference between Ω_{peak} and Ω_{null} in radians per second
δf	difference between Ω_{peak} and Ω_{null} in hertz
δt_d	time-delay resolution
SNR_{loss}	signal-to-noise ratio loss
Δt_{rw}	time duration associated with a range window
B_F	low-pass filter bandwidth
$Y(\omega)$	spectrum of the sampled signal
$y(n)$	sampled received signal
n	sample index
N	number of samples collected from a single point scatterer
$Y_M(\omega)$	magnitude of the spectrum
Δr	range delay relative to the center of the range window
$\delta\omega$	Rayleigh resolution in radians per sample
k	discrete Fourier transform (DFT) bin index
$Y(k)$	discrete Fourier transform
M	discrete Fourier transform size
N_{noise}	number of samples containing thermal noise collected over the receive window
SNR_β	signal-to-noise ratio using a filter with bandwidth β

SNR_{B_F}	signal-to-noise ratio at the output of a filter with bandwidth B_F
SNR_{DFT}	signal-to-noise ratio at the output of the DFT
f	frequency in hertz
f'_b	beat frequency including Doppler shift
t'_d	time delay offset associated with range-Doppler coupling

2.1.3.3 Stepped Chirp Waveforms

Variables associated with stepped chirp waveforms are as follows:

β	LFM intrapulse swept bandwidth
N_{sc}	number of pulses comprising a stepped chirp waveform
n	pulse index
$x_{tx}(t)$	transmit waveform
β	single-pulse bandwidth
β_{sc}	stepped chirp waveform's composite bandwidth
$x_{rcv}(t)$	received waveform
$LO_{rcv}(t)$	local oscillator signal applied to received waveform
$x_{BB}(t)$	received waveform mixed to baseband
p	new pulse index
$x_{BB}(m)$	sampled received waveform
m	sample index
T_c	time interval supporting the pulse width and receive window
$y_n(m)$	samples collected from the n -th pulse
$y'_n(m)$	interpolated signal associated with the n -th pulse
$z_n(m)$	output of the digital mixer operation associated with the n -th pulse
ϕ_n	phase correction applied to the n -th pulse
$z(m)$	frequency shifted, phase corrected, and time-aligned stepped chirp waveform
$X_n(\omega)$	DTFT of the n -th received pulse
$X(\omega)$	DTFT of a sampled baseband LFM waveform
$X^*(\omega)$	spectrum of the matched filter
$Y_n(\omega)$	spectrum of the n -th pulse having applied a matched filter on receive
δf_{DFT}	DFT bin size
k	DFT bin index
K	length of DFT
P	an integer

2.1.3.4 Nonlinear Frequency Modulated Waveforms

Variables associated with NFLM waveforms are as follows:

$x(t)$	notional waveform
$a(t)$	waveform's time-domain amplitude response
$\phi(t)$	waveform's time-domain phase response
$X(\Omega)$	waveform's spectrum
$ X(\Omega) $	magnitude of the spectrum
$\theta(\Omega)$	spectrum phase
β	bandwidth over which the waveform's frequency is swept
$W(\Omega)$	cosine on a pedestal weighting function defined in the frequency domain
h	parameter associated with cosine on a pedestal tapers

$W_{Taylor}(\Omega)$	Taylor weighting function defined in the frequency domain
F_m	Taylor coefficients
m	Taylor coefficient index
\bar{n}	n-bar used to define a Taylor weighting function
PSR	peak sidelobe ratio
a_0	average term in Fourier series
b_k	Fourier series coefficient for even signals
d_k	Fourier series coefficient for odd signals
Ω_0	fundamental frequency associated with a periodic signal

2.1.3.5 Stepped Frequency Waveforms

Variables associated with SF waveforms as follows:

N	number of pulses
R_0	range to a stationary point target
θ	measured phase
n	pulse index
$\Delta\theta$	phase difference between two pulses
$x(n)$	sample collected from the n -th pulse
$X(\omega)$	DTFT of sampled returns
ω_{R_0}	frequency in radians per sample; corresponds to the target's range
R	range
$\delta\omega$	Rayleigh resolution in radians per sample
ω_k	k -th discrete frequency
R_k	k -th discrete range
$X(k)$	discrete Fourier transform
k	DFT bin index
M	size of DFT
R_{gate}	location of range gate
L	physical length of a target
R_A	ambiguous range
$\hat{x}(n)$	samples containing a Doppler shift
r_{shift}	displacement in range due to a Doppler shift
\bar{r}_{shift}	normalized range displacement due to a Doppler shift
r_{spread}	spread in range due to a Doppler shift
\bar{r}_{spread}	normalized range spread due to a Doppler shift
v	radial velocity
\hat{v}	estimate of radial velocity
$x_{correct}$	correction factor applied to compensate for a Doppler shift

2.1.3.6 Quadriphase Codes

Variables associated with quadriphase codes are as follows:

c_n	biphase code indexed by n
q_n	quadriphase code generated from parent biphase code
s	a variable having a value of 1 or -1
N	length of biphase code
$p(t)$	subpulse envelope
τ_c	subpulse width of a biphase code

$z(t)$	complex signal formed by the quadriphase transformation
$a(t)$	envelope or magnitude of $z(t)$
$\phi(t)$	phase of $z(t)$
$y(t)$	quadriphase code transmitted by a radar and centered at baseband
$P(\Omega)$	spectrum of the half-cosine subpulse
$m(t)$	autocorrelation of half-cosine subpulse
QBR	ratio of quadriphase to biphasic peak sidelobe
a	peak sidelobe of biphasic code

2.1.3.7 Mismatched Filters

Variables associated with mismatched filters are as follows:

c_k	elements of a biphasic or polyphase code indexed by k
k	phase code element index
K	length of the phase code
z_m	coefficients associated with an M length finite impulse response filter
m	filter coefficient index
M	length of filter
y_n	output having applied a mismatched filter to a phase code
n	n -th output sample index
d_n	desired mismatch filtered response
e_n	error signal or difference between desired response and actual response
E	sum of the squared error
y	column vector containing the filtered response
C	matrix containing shifted copies of the phase code
z	column vector containing the mismatched filter
H	Hermitian operator
d	column vector containing the desired response
W	weighting matrix
w	an element of the weighting matrix

2.1.4 Acronyms

Acronyms used in this chapter are as follows:

ADC	analog-to-digital converter
BTQ	biphasic-to-quadriphase
COHO	coherent oscillator
CW	continuous wave
DFT	discrete Fourier transform
DTFT	discrete time Fourier transform
EMI	electromagnetic interference
ENOB	effective number of bits
FD	frequency domain
FFT	fast Fourier transform
FMCW	frequency modulated continuous wave
FSK	frequency shift keying
HRR	high-resolution range
IF	intermediate frequency

ISR	integrated sidelobe ratio
LFM	linear frequency modulated
LPG	loss in processing gain
LS	least squares
MF	matched filter
MISR	minimum integrated sidelobe ratio
MMF	mismatched filter
MPS	minimum peak sidelobe
MSK	minimum shift keying
NLFM	nonlinear frequency modulated
PC	pulse compression
PRF	pulse repetition frequency
PRI	pulse repetition interval
PSP	principle of stationary phase
PSR	peak sidelobe ratio
RF	radio frequency
SAR	synthetic aperture radar
SNR	signal-to-noise ratio
SF	stepped frequency
TD	time domain
WLS	weighted least squares

2.2 | STRETCH PROCESSING

2.2.1 Introduction

Radar systems employ fine resolution when forming a target's range profile. These systems are, by nature, wideband given that range resolution is inversely proportional to waveform bandwidth [1]. Resolutions on the order of a half foot or less are common, requiring a 1 GHz or greater bandwidth. Designing and building hardware to support instantaneous (i.e., intrapulse) wideband waveforms is challenging.

Stretch processing [2–6] is a technique applied to an LFM waveform [1] to reduce the bandwidth required to process the signal (i.e., processing bandwidth) while maintaining the range resolution afforded by the waveform's transmit bandwidth. The system's processing bandwidth is defined by the ADC's sampling rate. SARs and other fine range resolution systems often employ stretch processing.

2.2.2 Processing Bandwidth

ADCs are required to sample at a rate equal to or greater than the waveform bandwidth to satisfy the Nyquist criteria, and even higher rates are needed to support direct intermediate frequency (IF) sampling [36]. Technology places constraints on the achievable sampling rate and effective number of bits (ENOB) [37–39]. As the sampling rate is increased, the trend is toward lower ENOB. An ADC's ENOB affects the radar's dynamic range and quantization noise. Thus, both the ADC's sampling rate and ENOB are considered when selecting a device.

The data bus and signal processor's requirements are driven by the data rate and the number of samples collected per pulse. The minimum number of samples is equal to the waveform's time–bandwidth product. For pulse compression waveforms, time–bandwidth products range from 2 to 1,000,000 or even greater. For an LFM waveform with a $10\ \mu\text{sec}$ pulse width and a 1 GHz bandwidth, the minimum number of samples to process per pulse is 10,000. The data bus and signal processor must be sized to transfer and process large quantities of data in real time, which may be a challenge. Stretch processing reduces both the data rate and number of samples to process while maintaining the range resolution afforded by the transmit bandwidth.

2.2.3 Technique Overview

Caputi [2] developed the stretch processing technique, which converts a received LFM pulse into a tone whose frequency is fixed and is proportional to the target's time delay. A filter with a passband less than the waveform's bandwidth is used to limit the range of frequencies passed to the ADC. Limiting the processing bandwidth constrains the range of observable time delays or, equivalently, the size of the range window. In general, the constraint or limitation placed on the range window is acceptable when there is knowledge of the distance to the target or area to be imaged. For example, a missile defense radar may employ stretch processing to image a target that is already under track. In a SAR system, the size and location of the area to be imaged is known. The samples at the output of the ADC are processed using a discrete Fourier transform (DFT) that compresses the returns in range. The range resolution at the output of the DFT is inversely proportional to the transmit bandwidth and is not limited by the reduced ADC sampling rate.

2.2.4 Implementation

Stretch processing is uniquely associated with an LFM waveform and is implemented in both pulsed and frequency modulated continuous wave (FMCW) radar systems. In this chapter, the focus is placed on pulsed systems. A modern implementation consists of the following:

1. The transmission of a wideband, linear frequency modulated waveform
2. A mixing operation on receive that converts the received waveform into a set of tones with frequencies proportional to the relative time delay between scatterers
3. A filter following the mixer to limit the highest frequency passed to the ADC
4. An ADC sampling at a rate proportional to the filter bandwidth
5. A DFT to compress the received waveform

Each step is examined in the following sections.

2.2.4.1 Transmit Waveform

Consider an LFM waveform, centered at an IF, f_1 ,

$$x(t) = \cos \left(2\pi f_1 t + \pi \frac{\beta}{\tau} \left(t - \frac{\tau}{2} \right)^2 + \theta_1 \right) \quad 0 \leq t \leq \tau \quad (2.1)$$

where β is the bandwidth over which the waveform is swept, τ is the pulse length, and θ_1 is the phase of the oscillator used to generate f_1 . Approaches for synthesizing an

LFM waveform include the use of a voltage-controlled oscillator or a digital waveform generator. In this example, a heterodyne transmit/receive architecture with two mixer stages is employed. A homodyne (single-stage) receiver or a heterodyne receiver with more than two stages may also be used.

On transmit, the IF signal is mixed to an RF using a second oscillator, which is

$$LO_{2,tx} = \cos(2\pi f_2 t + \theta_2) \quad t \geq 0 \quad (2.2)$$

and is mixed with the signal in equation (1) to produce

$$x(t) = \cos\left(2\pi(f_1 + f_2)t + \pi\frac{\beta}{\tau}\left(t - \frac{\tau}{2}\right)^2 + \theta_1 + \theta_2\right) \quad 0 \leq t \leq \tau \quad (2.3)$$

with the lower sideband removed via filtering and the result scaled to unit amplitude. The signal is now centered at an RF $f_0 = f_1 + f_2$. The transmit signal propagates through space and is reflected by targets and clutter.

2.2.4.2 Receiver

In a pulsed system, the receive action is timed to coincide with a specific time delay t_{rcv} , which is referenced to the center of the range window. Signals reflected toward the radar are mixed on receive with a copy of the transmit modulation. Consider the return from a point target located at time delay t_d

$$x_r(t) = \cos\left(2\pi(f_1 + f_2)(t - t_d) + \pi\frac{\beta}{\tau}\left(\left(t - \frac{\tau}{2}\right) - t_d\right)^2 + \theta_1 + \theta_2\right) \quad t_d \leq t \leq t_d + \tau \quad (2.4)$$

The first receive oscillator consists of an intermediate frequency, f_2 , and a linear frequency modulated component that matches the transmit waveform

$$LO_{1,rcv} = \cos\left(2\pi f_2 t + \pi\frac{\beta}{\tau}\left(\left(t - \frac{\tau}{2}\right) - t_{rcv}\right)^2 + \theta_2\right) \quad t_{rcv} \leq t \leq t_{rcv} + \tau \quad (2.5)$$

The output of the first mixer is

$$y(t) = \cos\left(2\pi f_1(t - t_d) - 2\pi f_2 t_d + \pi\frac{\beta}{\tau}\left(-2\left(t - \frac{\tau}{2}\right)(t_d - t_{rcv}) + (t_d^2 - t_{rcv}^2)\right) + \theta_1\right) \quad (2.6)$$

where $t_{rcv} \leq t \leq (t_{rcv} + \tau - t_d)$ for a time delay less than or equal to the center of the receive window, (i.e., $t_{rcv} - \tau \leq t_d \leq t_{rcv}$), and $t_d \leq t \leq (t_{rcv} + \tau)$ for a time delay greater than or equal to the center of the receive window, (i.e., $t_{rcv} \leq t_d \leq t_{rcv} + \tau$). The upper sideband associated with the mixer output is removed via filtering, and the signal is centered at an IF f_1 . Note that the amplitude of the return, which is dependent on the target's RCS and radar parameters, is suppressed to simplify the expressions. The stretch processor is linear; thus, the following analysis applies equally to a superposition of returns from targets at different delays.

The signal and timing relationships are illustrated in Figure 2-1. The receive oscillator, in Figure 2-1, is centered at RF for illustrative purposes. The homodyne representation is easier to depict and yields equivalent baseband results. The receive oscillator is turned on at time t_{rcv} . The returns from two targets are present at time delays t_{d1} and t_{d2} , respectively.

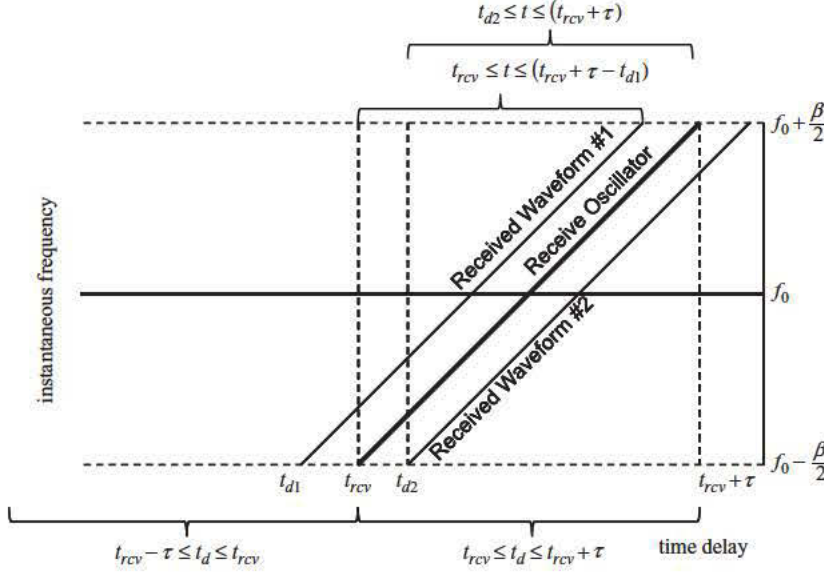


FIGURE 2-1 ■ With stretch processing, the radar mixes received waveforms with a copy of the transmit modulation. The receive oscillator is turned on at a time delay corresponding to the center of the range window.

The signal in equation (2.6) is moved to baseband by mixing it with a second receive oscillator

$$LO_{2,rcv} = \cos(2\pi f_1 t + \theta_1) \quad t \geq 0 \quad (2.7)$$

The mixing operation produces

$$y(t) = \cos \left(-2\pi (f_1 + f_2) t_d + \pi \frac{\beta}{\tau} \left(-2 \left(t - \frac{\tau}{2} \right) (t_d - t_{rcv}) + (t_d^2 - t_{rcv}^2) \right) \right) \quad (2.8)$$

where $t_{rcv} \leq t \leq (t_{rcv} + \tau - t_d)$ for $t_{rcv} - \tau \leq t_d \leq t_{rcv}$ and $t_d \leq t \leq t_{rcv} + \tau$ for $t_{rcv} \leq t_d \leq t_{rcv} + \tau$.

To simplify the expression, let $\Delta t_d = t_d - t_{rcv}$ and $\varphi = \pi \beta / \tau (t_d^2 - t_{rcv}^2)$. The quadratic phase, φ , is known as the residual video phase (RVP) and is a potential source of blurring in SAR imagery [40]. Making the proposed substitutions, equation (2.8) reduces to

$$y(t) = \cos \left(-2\pi (f_1 + f_2) t_d + \pi \frac{\beta}{\tau} \left(-2 \left(t - \frac{\tau}{2} \right) \Delta t_d \right) + \varphi \right) \quad (2.9)$$

Arranging terms yields

$$y(t) = \cos \left(-2\pi \frac{\beta}{\tau} \Delta t_d t - 2\pi (f_1 + f_2) t_d + 2\pi \frac{\beta}{2} \Delta t_d + \varphi \right) \quad (2.10)$$

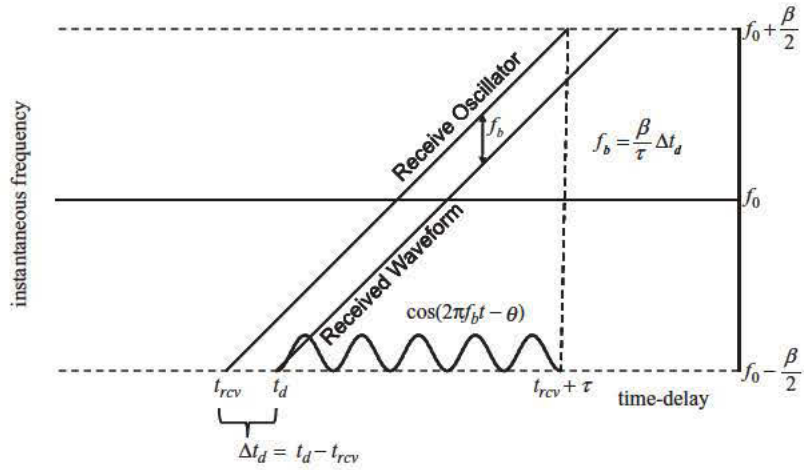
Grouping the constant phase terms, let $\theta = -2\pi (f_1 + f_2) t_d + 2\pi \beta / 2 \Delta t_d + \varphi$. Equation (2.10) reduces to

$$y(t) = \cos \left(2\pi \frac{\beta}{\tau} \Delta t_d t - \theta \right) \quad (2.11)$$

where $t_{rcv} \leq t \leq (t_{rcv} + \tau - t_d)$ for $(t_{rcv} - \tau) \leq t_d \leq t_{rcv}$ and $t_d \leq t \leq (t_{rcv} + \tau)$ for $t_{rcv} \leq t_d \leq (t_{rcv} + \tau)$. The change of the argument's sign in equation (2.11) is valid since

FIGURE 2-2 ■

Mixing the receive signal with the local oscillator produces a tone or beat frequency proportional to the relative time delay between the target and the center of the range window.



the cosine is an even function. Mixing the received signal with a copy of the transmit modulation produces a sinusoid in equation (2.11) with beat frequency

$$f_b = \frac{\beta}{\tau} \Delta t_d \quad (2.12)$$

As illustrated in Figure 2-2, the beat frequency represents the difference in instantaneous frequency between the oscillator ramp and the received signal. The process of converting an LFM waveform into a sinusoid via the mixing operation is referred to as *deramping*.

The processor measures time relative to the center of the range window. One may compute absolute time delay by adding the time delay associated with center of the range window, as measured via the system's reference clock, to the time delay measured by the stretch processor. In this architecture, the receiver is designed to limit the processed time delays such that $\Delta t_d \ll \tau$ and thus $f_b \ll \beta$. The approach and reason for limiting the processed time delay is discussed in Section 2.2.4.8.

2.2.4.3 Processor Architecture

The receiver and signal processor components are depicted in Figure 2-3. The beat frequency is generated in the first mixer stage and includes a frequency offset, f_1 . The signal is then mixed to baseband using a coherent detector. The coherent detector, consisting of in-phase and quadrature channels, provides a means for distinguishing returns from scatterers located at time delays less than t_{rcv} from those located at time delays greater than t_{rcv} . The analytical representation of the baseband signal is

$$y(t) = \exp \left(j \left(2\pi \frac{\beta}{\tau} \Delta t_d t - \theta \right) \right) \quad (2.13)$$

where $t_{rcv} \leq t \leq (t_{rcv} + \tau - t_d)$ for $t_{rcv} - \tau \leq t_d \leq t_{rcv}$ and $t_d \leq t \leq t_{rcv} + \tau$ for $t_{rcv} \leq t_d \leq t_{rcv} + \tau$.

The low-pass filter preceding the ADC limits the maximum beat frequency, and the basis for selecting the filter's bandwidth is presented in Section 2.2.4.8. The ADC is required to sample at a rate equal to or greater than the filter bandwidth. To address the filter's roll-off at the band edges, a sampling rate slightly greater than the filter bandwidth may be selected. Sampled returns are compressed in range by applying a DFT to the signal. An amplitude weighting may also be applied, prior to the DFT, to reduce the range sidelobes.

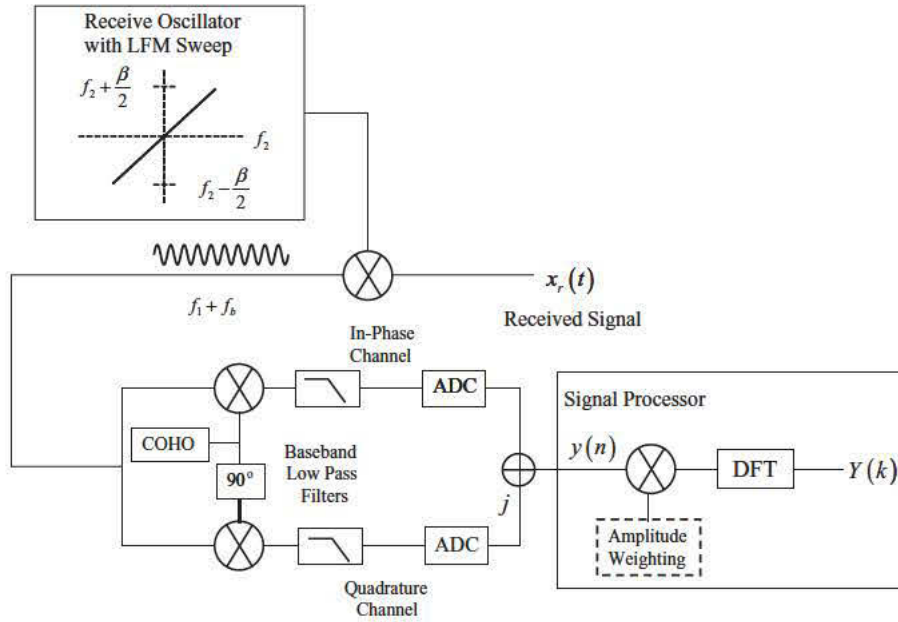


FIGURE 2-3 ■ The stretch processor consists of a deramp operation, a low-pass filter to limit the maximum beat frequency, and a DFT to compress the returns in range.

2.2.4.4 Spectrum of the Baseband Signal

Before examining the signal at the output of the DFT, let's begin by applying the Fourier transform to the continuous-time, complex sinusoid in equation (2.13), which represents the return from a point target after the deramp operation. The spectrum of the signal provides valuable insight into principles associated with the stretch processor. The results will be analogous to the sampled returns processed using a DFT.

Consider the case where $t_{rcv} - \tau \leq t_d \leq t_{rcv}$. The Fourier transform of the signal in equation (2.13) is

$$Y(\Omega) = \int_{t_{rcv}}^{t_d + \tau} \exp\left(j\left(2\pi\frac{\beta}{\tau}\Delta t_d t - \theta\right)\right) \exp(-j\Omega t) dt \quad (2.14)$$

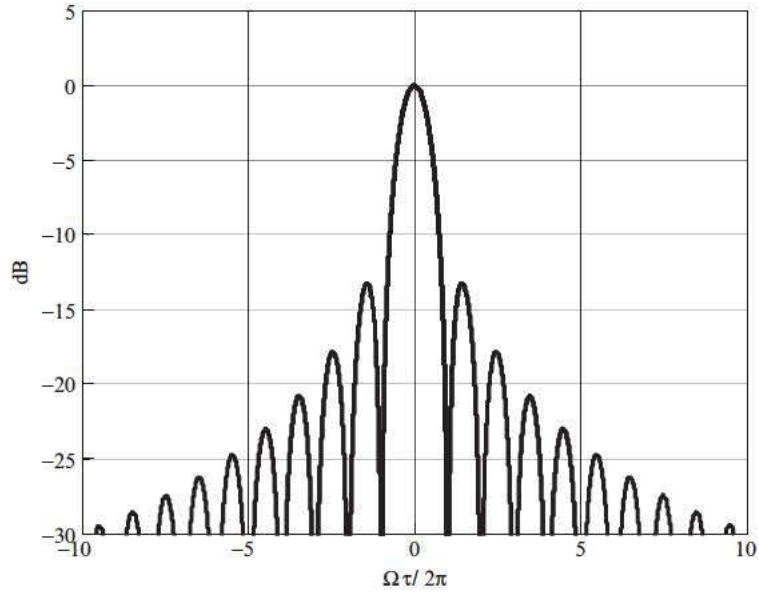
Evaluating the integral yields

$$Y(\Omega) = \frac{\exp\left(j\left(2\pi\frac{\beta}{\tau}\Delta t_d - \Omega\right)t\right)}{j\left(2\pi\frac{\beta}{\tau}\Delta t_d - \Omega\right)} \Bigg|_{t_{rcv}}^{t_d + \tau} \exp(-j\theta) \quad (2.15)$$

and for $t_{rcv} \leq t_d \leq t_{rcv} + \tau$

$$Y(\Omega) = \frac{\exp\left(j\left(2\pi\frac{\beta}{\tau}\Delta t_d - \Omega\right)t\right)}{j\left(2\pi\frac{\beta}{\tau}\Delta t_d - \Omega\right)} \Bigg|_{t_d}^{t_{rcv} + \tau} \exp(-j\theta) \quad (2.16)$$

FIGURE 2-4 ■ The stretch processor produces a sinc response at the output of the DFT.



The expressions in equations (2.15) and (2.16) reduce to

$$Y_M(\Omega) = \left| \frac{d \sin \left(\left(2\pi \frac{\beta}{\tau} \Delta t_d - \Omega \right) d/2 \right)}{\left(2\pi \frac{\beta}{\tau} \Delta t_d - \Omega \right) d/2} \right| \quad (2.17)$$

where $d = -\tau \left(1 - \frac{|\Delta t_d|}{\tau} \right)$ and $0 \leq |\Delta t_d| \leq \tau$. Equation (2.17) contains a sinc function and represents the range compressed response associated with a scatterer located at time delay t_d . A plot of the compressed response is presented in Figure 2-4 with the target located in the center of the range window ($\Delta t_d = 0$). The response consists of a main lobe and a sidelobe structure. Range resolution is proportional to the width of the main lobe, and the peak sidelobes occur adjacent to the main lobe and are 13.2 dB below the peak of the response.

2.2.4.5 Range Resolution

The Rayleigh criterion [1] is often used to quantify range resolution and is defined as the distance between the peak and first null of the compressed response. The response in equation (2.17) has a peak value at $\Omega_{peak} = 2\pi \beta / \tau \Delta t_d$, and the first null, nearest the peak, occurs when

$$\frac{\left(2\pi \frac{\beta}{\tau} \Delta t_d - \Omega_{null} \right) d}{2} = \pm \pi \quad (2.18)$$

or

$$\Omega_{null} = \pm \frac{2\pi}{d} + 2\pi \frac{\beta}{\tau} \Delta t_d \quad (2.19)$$

The separation between the peak and first null is

$$\delta\Omega = |\Omega_{peak} - \Omega_{null}| = \frac{2\pi}{|d|} \quad (2.20)$$

Equation (2.20) represents the Rayleigh resolution in radians per second, and the resolution in hertz is

$$\delta f = \frac{1}{|d|} \quad (2.21)$$

Consider two point targets separated in time by δt_d . The corresponding difference in beat frequencies is $\delta t_d \beta / \tau$. Setting the frequency resolution achieved by the compressed response in equation (2.17) equal to the difference in beat frequencies

$$\frac{1}{|d|} = \frac{\beta}{\tau} \delta t_d \quad (2.22)$$

and solving for the time-delay resolution

$$\delta t_d = \frac{1}{\beta} \frac{1}{\left(1 - \frac{|\Delta t_d|}{\tau}\right)} \quad (2.23)$$

or in terms of range resolution is

$$\delta R = \frac{c}{2\beta} \frac{1}{\left(1 - \frac{|\Delta t_d|}{\tau}\right)} \quad (2.24)$$

where c is the speed of light. Now, consider a scatterer located at a time delay equal to the center of the range window (i.e., $t_d = t_{rcv}$ or $\Delta t_d = 0$). The resultant range resolution is

$$\delta R_{\Delta t_d=0} = \frac{c}{2\beta} \quad (2.25)$$

For scatterers located at time delays less than or greater than t_{rcv} , range resolution degrades by a factor of $1/(1 - |\Delta t_d|/\tau)$ as shown in equation (2.24). The loss in resolution occurs because the receive oscillator extends only over a time interval equal to the original pulse length. A return at a range other than the center of the range window mixes with the oscillator, producing a tone whose duration is less than the original pulse length. The shortened sinusoid achieves less frequency resolution and correspondingly less range resolution.

2.2.4.6 SNR Loss

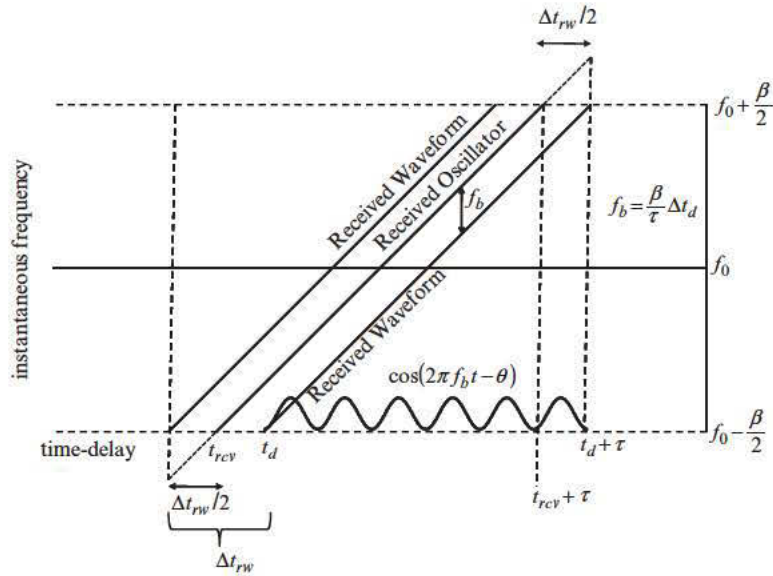
In addition to the degradation in resolution, a loss in SNR occurs when the sinusoid (or mixing product) is shortened in time:

$$SNR_{Loss} = 1 - \frac{|\Delta t_d|}{\tau} \quad 0 \leq |\Delta t_d| < \tau \quad (2.26)$$

2.2.4.7 Oscillator Over Sweep

The loss in SNR and degradation in resolution may be eliminated by extending the oscillator sweep to accommodate the size of the range window, as illustrated in Figure 2-5.

FIGURE 2-5 ■ A stretch processor employs an over sweep to account for the size of the range window.



Consider a range window, ΔR , with corresponding time extent, Δt_{rw} , centered about t_{rcv} . To preserve the mixer product, the ends of the oscillator sweep are extended by $\Delta t_{rw}/2$. The mixer product is shown starting at time delay t_d and ending at time delay $t_d + \tau$. A scatterer located anywhere within the range window will produce a mixer product of duration τ and thus will achieve the finest range resolution and maximum SNR afforded by the transmit waveform.

2.2.4.8 Filter Bandwidth

Within the receiver, a low-pass filter precedes the ADC and serves to limit the range of beat frequencies and thus the size of the range window. To reduce the requirements placed on the ADC, data bus, and signal processor, the filter bandwidth is intentionally chosen to be less than the bandwidth of the LFM waveform. Consider a range window of extent ΔR . The filter bandwidth required to pass beat frequencies associated with scatterers located within the range window is

$$B_F = \frac{\beta}{\tau} \frac{2\Delta R}{c} \quad (2.27)$$

In general, the pulse width and range window are chosen such that $2\Delta R/c \ll \tau$ and as a result $B_F \ll \beta$. The ADC samples at a rate, F_s , where

$$F_s \geq B_F \quad (2.28)$$

The real advantage associated with stretch processing is the relationship $F_s \ll \beta$. Stretch processing allows range window extent to be traded for a reduced sampling rate (i.e., processing bandwidth) while achieving a range resolution afforded by the transmit bandwidth.

2.2.4.9 Compressed Range Response

A high-resolution range (HRR) profile is created by computing the power spectrum associated with the complex samples, $y(n)$, at the output of the ADC in Figure 2-3. The DFT

is applied in the signal processor to compute the spectrum and is a sampled version of the discrete time Fourier transform (DTFT),

$$Y(\omega) = \sum_{n=0}^{N-1} y(n) \exp(-jn\omega) \quad (2.29)$$

where ω is continuous. Sampling the received signal in equation (2.13) produces

$$y(n) = \exp\left(j2\pi \frac{\beta}{\tau} \Delta t_d (nT_s)\right) \exp(-j\theta) \quad n = 0, 1, \dots, (N-1) \quad (2.30)$$

where T_s is the ADC sampling period, and $F_s = 1/T_s$. If no oversweep is employed, the number of samples, N , collected from a specific scatterer is a function of the scatterer's relative position within the receive window

$$N = \left\lceil F_s \tau \left(1 - \frac{|\Delta t_d|}{\tau}\right) \right\rceil \quad (2.31)$$

and if an oversweep of the oscillator is employed, the number of samples is independent of time delay

$$N = \lceil F_s \tau \rceil \quad (2.32)$$

Applying the DTFT to the sampled signal in equation (2.30), the magnitude of the response is

$$Y_M(\omega) = \left| \frac{\sin\left(\frac{N}{2} \left(\omega - \frac{2\pi \beta \Delta t_d}{F_s \tau}\right)\right)}{\sin\left(\frac{1}{2} \left(\omega - \frac{2\pi \beta \Delta t_d}{F_s \tau}\right)\right)} \right| \quad (2.33)$$

The response in equation (2.33) is known as a digital sinc function or Dirichlet function. The spectrum is periodic in 2π , and a single period is defined over $-\pi \leq \omega < \pi$. The shape of the response is very similar to that in equation (2.17) and matches the shape of the main lobe and near-in sidelobes. The conversion from digital frequency, ω , to relative time delay is

$$\Delta t_d = \omega \frac{F_s \tau}{2\pi \beta} \quad (2.34)$$

and the conversion to relative range delay is

$$\Delta r = \omega \frac{F_s \tau c}{2\pi \beta 2} \quad (2.35)$$

The achieved range resolution is a function of the main lobe width in equation (2.33). If the relative time delay is zero, the peak occurs at $\omega = 0$, and the first null occurs at $N/2\omega = \pi$. The Rayleigh resolution [1] is then

$$\delta\omega = \frac{2\pi}{N} \quad (2.36)$$

Range resolution is related to frequency resolution via

$$\delta R = \frac{2\pi}{N} \frac{F_s}{2\pi} \frac{\tau}{\beta} \frac{c}{2} \quad (2.37)$$

or

$$\delta R = \frac{F_s \tau}{\left[F_s \tau \left(1 - \frac{\Delta t_d}{\tau} \right) \right]} \frac{c}{2\beta} \quad (2.38)$$

If an oversweep is employed, the resulting range resolution is

$$\delta R = \frac{F_s \tau}{[F_s \tau]} \frac{c}{2\beta} \quad (2.39)$$

If $F_s \tau$ is an integer, then δR is exactly equal to $c/2\beta$.

2.2.4.10 Discrete Fourier Transform

A DTFT may not be implemented in a signal processor given that ω is continuous, but a DFT may be applied and is efficiently implemented using a fast Fourier transform (FFT) [41]. The DFT samples the DTFT response at equally spaced frequencies defined by

$$\omega_k = 2\pi \frac{nk}{M} \quad k = 0, 1, \dots, (M-1) \quad (2.40)$$

where $M \geq N$. The DFT is defined as

$$Y(k) = \sum_{n=0}^{N-1} y(n) \exp\left(-j2\pi \frac{nk}{M}\right) \quad k = 0, 1, \dots, (M-1) \quad (2.41)$$

For $M > N$, zero padding is assumed.

The DFT of the signal in equation (2.30), which may be obtained by sampling the DTFT in equation (2.33), is

$$|Y(k)| = \left| \frac{\sin\left(\frac{N}{2} \left(\frac{2\pi k}{M} - \frac{2\pi \beta \Delta t_d}{F_s \tau} \right)\right)}{\sin\left(\frac{1}{2} \left(\frac{2\pi k}{M} - \frac{2\pi \beta \Delta t_d}{F_s \tau} \right)\right)} \right| \quad (2.42)$$

Time delay is related to the k -th DFT bin via

$$\Delta t_d = \frac{\tau}{\beta} F_s \frac{k}{M} \quad k = 0, 1, \dots, (M-1) \quad (2.43)$$

Note that increasing M decreases the DFT bin spacing resulting in a decrease in straddle loss [1]. Range resolution is governed by the continuous spectrum in equation (2.33) and not the sampled response associated with the DFT in equation (2.42).

2.2.5 Example System Parameters

Table 2-1 contains a comparison of stretch and correlation processing [1] for a notional radar employing a 1 GHz LFM waveform with a 100 μ sec pulse width. In both instances, the nominal range resolution is 0.15 meters.

TABLE 2-1 ■ A Comparison of the Parameter Values Associated with Correlation and Stretch Processing

Parameter	Correlation Processing	Stretch Processing
Pulse width (μsec)	100	100
Waveform bandwidth (MHz)	1000	1000
Range resolution (m)	0.15	0.15
Baseband filter bandwidth (MHz)	1000	100
Processing bandwidth (MHz)	1000	100
Processing time–bandwidth product (minimum)	100,000	10,000
Range window (km)	all range	1.5

A correlator supports the processing of all ranges. The minimum filter bandwidth is based on the waveform bandwidth and in this example is 1 GHz. The minimum number of samples to process is equal to the product of the sampling rate and the pulse width and is 100,000.

With the stretch processor, the filter bandwidth is reduced here by a factor of 10 and is set at 100 MHz. The ADC sampling rate and processing bandwidth are also reduced, leading to a reduction in data rate and throughput requirements. The range window is then limited by the low-pass filter to 1.5 km. A reduction in processing bandwidth is achieved at a cost of limiting the size of the range window.

2.2.6 Processing Gain

Stretch processing achieves a pulse compression gain equal to the waveform time–bandwidth product. To prove this, consider a stretch processor where the sampling rate is equal to the filter bandwidth. When an oversweep is applied, the complex tone associated with a point target consists of N samples defined by

$$N = [B_F \tau] \quad (2.44)$$

The number of noise samples collected over a processing window is

$$N_{noise} = B_F \left[\tau + \frac{2\Delta R}{c} \right] \quad (2.45)$$

A sampling rate greater than the filter bandwidth increases the number of samples, but the noise samples are correlated. Thus, the additional samples do not translate into an SNR gain.

Consider an LFM waveform with bandwidth β that is passed through a linear phase filter with the same bandwidth. The SNR at the output of the filter is denoted SNR_β . In the stretch processor, the deramp signal (mixer product) is passed through a linear phase filter with bandwidth B_F to limit the range of frequencies. The SNR at the output of the filter is

$$SNR_{B_F} = SNR_\beta \frac{\beta}{B_F} \quad (2.46)$$

Since $B_F < \beta$, the signal experiences an SNR gain.

The DFT provides additional gain by coherently processing the received signal. Coherent signals experience a power gain through the DFT equal to N^2 . In contrast, thermal noise experiences a power gain equal to N_{noise} . The SNR gain through the DFT is then N^2/N_{noise} . At the output of the DFT, the resultant SNR is

$$SNR_{DFT} = SNR_{\beta} \frac{\beta}{B_F} \frac{N^2}{N_{noise}} \quad (2.47)$$

With stretch processing, $2\Delta R/c \ll \tau$. Therefore, the number noise samples is approximately

$$N_{noise} \approx B_F \tau. \quad (2.48)$$

Substituting equations (2.44) and (2.48) into equation (2.47), we arrive at an expression for the SNR gain achieved by a signal having passed through the filter and DFT

$$SNR_{DFT} = SNR_{\beta} \frac{\beta}{B_F} B_F \tau = SNR_{\beta} \beta \tau \quad (2.49)$$

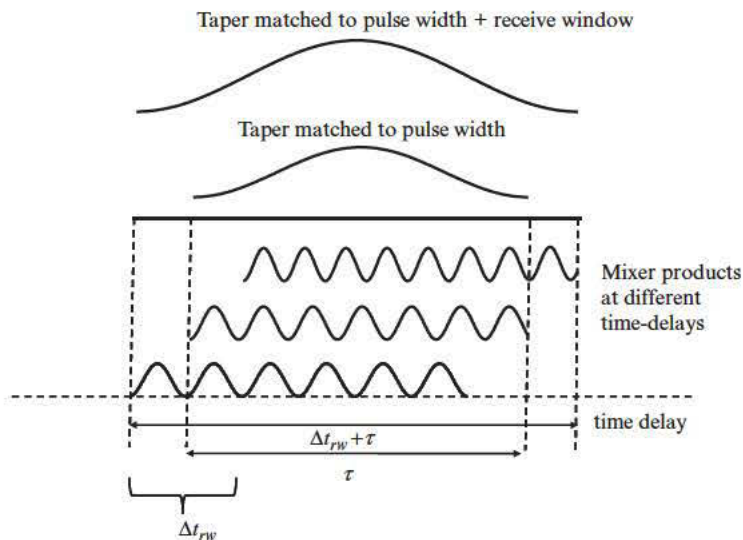
The pulse compression gain, $\beta \tau$, at the output of the stretch processor is equivalent to that achieved with correlation processing [1].

2.2.7 Range-Sidelobe Suppression

In radar, sidelobes may mask the presence of a smaller scatterer or degrade the overall quality of an image. Range sidelobes are reduced by applying an amplitude taper to the data prior to the DFT. As with correlation processing, the amplitude taper reduces the sidelobes at the expense of degraded resolution and a loss in SNR [1].

Scatterers located at different ranges produce returns that are not aligned in time and as a consequence experience an asymmetric weighting when a taper is applied. An asymmetric weighting degrades the sidelobe response. The length of the amplitude taper may be matched to the pulse width or the collection time (pulse width + range window)

FIGURE 2-6 ■ A time misalignment of the mixing products results in asymmetric weighting when a taper is applied.



extent) as illustrated in Figure 2-6. The degree of distortion in the compressed response is a function of two metrics: (1) the ratio of the range window extent to the pulse width (with smaller ratios resulting in the least amount of distortion); and (2) the alignment of the taper relative to a specific time delay. The distortion is thus time-delay dependent. For cases where the pulse width is much longer than the range window, the effects are often ignored.

Kellog [42] describes a technique for time aligning the deramped signals using a dispersive filter. The filter is applied prior to the weighting and has a group delay equal to

$$t_{gd} = -\frac{\tau}{\beta} f \quad (2.50)$$

where f is frequency in hertz.

2.2.8 Range-Doppler Coupling

The stretch processed LFM waveform exhibits range-Doppler coupling. Figure 2-7 shows the impact of Doppler shift on the beat frequency. The frequency separation between the local oscillator and the LFM return is modified by the Doppler shift, f_d . The resultant beat frequency is

$$f'_b = \frac{\beta t_d}{\tau} - f_d \quad (2.51)$$

and the additional time delay or offset due to the Doppler shift is

$$t'_d = -\frac{\tau}{\beta} f_d \quad (2.52)$$

The offset in equation (2.52) is equivalent to the range-Doppler coupling observed when correlation processing is applied [1]. To avoid eclipsing, the filter's bandwidth must be sized to accommodate the largest beat frequency plus the largest anticipated Doppler shift.

2.2.9 Summary

Stretch processing is implemented in modern, high-resolution radar systems to reduce processing requirements without sacrificing range resolution. The technique is applied to

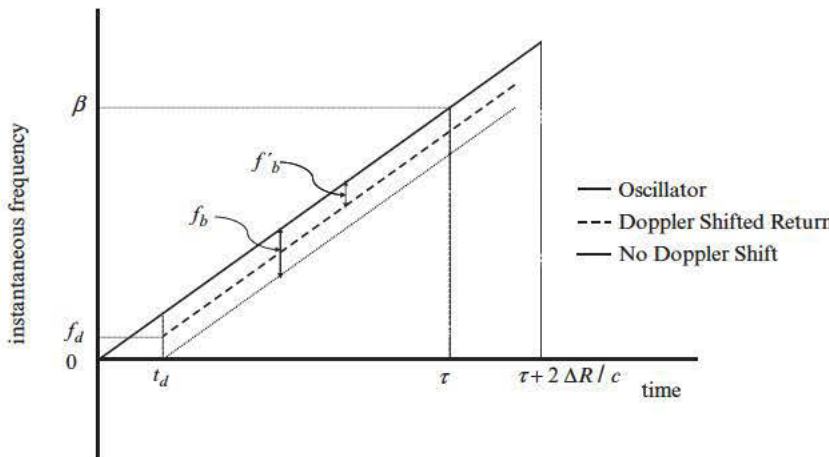


FIGURE 2-7 ■ The beat frequency is modified by a Doppler shift resulting in an error in the reported range to the target.

a wideband LFM waveform on receive and converts relative time delay into frequency. A filter limits the frequencies passed to the ADC and defines the processing bandwidth. The filter bandwidth is chosen to be less than the transmit bandwidth. The sampled waveform is compressed using a DFT, and the resultant range resolution is inversely proportional to the transmit bandwidth. The filter does limit the size of the range window to a fraction of the pulse width; however, in high-resolution systems the range to the target or imaged area is known with sufficient precision to position the window about the region of interest.

2.3 | STEPPED CHIRP WAVEFORMS

2.3.1 Introduction

As the name implies, a stepped chirp waveform [7–15] consists of LFM pulses that are stepped in frequency to reduce the waveform's instantaneous bandwidth. On transmit, the waveform's reduced intrapulse bandwidth lessens dispersion effects in some systems (e.g., a wideband, phased array radar that does not employ time delay units). On receive, the stepped chirp waveform supports a lower sampling rate, similar to stretch processing, but does not limit the size of the range window. The waveform is composed of N_{sc} LFM pulses, each with swept bandwidth, β , that are separated in time by the pulse repetition interval (PRI) and that partially overlap or are contiguous in frequency. Successive pulses are separated in frequency by a constant offset, Δf . In practice, the number of pulses used to construct the waveform is relatively small (e.g., $1 < N_{sc} \leq 10$).

On receive, each pulse is mixed to baseband or to a common intermediate frequency and sampled a rate sufficient to support the intrapulse bandwidth. The composite waveform is reconstructed in the signal processor by shifting the returns by the appropriate frequency offset and stitching them together to form a coherent wideband signal. The bandwidth of the composite waveform, β_{sc} , is $\beta + (N_{sc} - 1) \Delta f$. The requirements placed on the ADC sample rate are driven by the single pulse bandwidth, β , and the range window extent is limited by the PRI and not the low pass filter as in the case of stretch processing. As expected, the waveform's reconstruction does increase the computational requirements levied on the signal processor. With a stepped chirp waveform, the pulse repetition frequency (PRF) must be increased by a factor of N_{sc} to maintain the effective rate required to support multiple-pulse, coherent processing (e.g., SAR or Doppler processing) [10].

2.3.2 Transmit Waveform

The stepped chirp waveform consists of N_{sc} LFM pulses, each with swept bandwidth β , that are shifted in frequency from pulse-to-pulse by a constant frequency offset Δf where $\Delta f \leq \beta$ as illustrated in Figure 2-8. The n -th transmit pulse is modeled as

$$x_{tx}(t) = \cos \left(2\pi (f_0 + (-(N_{sc} - 1)/2 + n) \Delta f) t + \pi \frac{\beta}{\tau} \left(t - \frac{\tau}{2} \right)^2 \right) \quad (2.53)$$

where $0 \leq t \leq \tau$, and $0 \leq n \leq (N_{sc} - 1)$, and f_0 is the transmit center frequency. Time in equation (2.53) is referenced to the beginning of each pulse, but the individual pulses are separated by the PRI, T , where $T > \tau$. In the text, an odd number of pulses is used to develop a mathematical model, but a similar analysis applies for an even number of pulses.

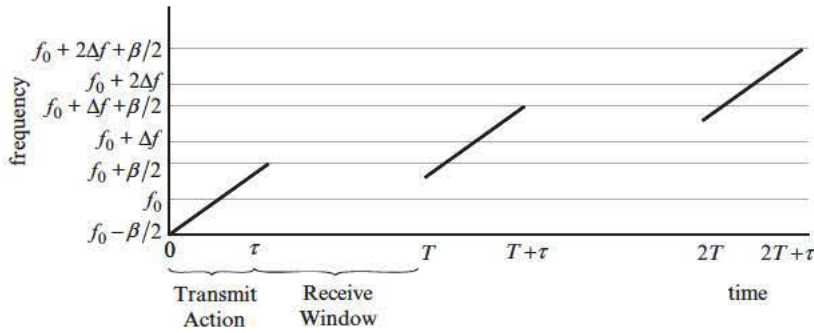


FIGURE 2-8 ■ A notional stepped chirp waveform consisting of three pulses and a frequency step size Δf .

2.3.3 Received Waveform

Consider a point target located at time delay t_d . After transmitting the pulse in equation (2.53), the received pulse (at RF) is

$$x_{rcv}(t) = \cos \left(2\pi (f_0 + (-(N_{sc} - 1)/2 + n) \Delta f) (t - t_d) + \pi \frac{\beta}{\tau} \left((t - t_d) - \frac{\tau}{2} \right)^2 \right) \quad (2.54)$$

where $t_d \leq t \leq \tau + t_d$ and $0 \leq n \leq (N_{sc} - 1)$. Within the receiver, the signal is mixed to baseband or to a common IF. The local oscillator used to mix the signal to baseband is

$$LO_{rcv}(t) = \cos(2\pi(f_0 + (-(N_{sc} - 1)/2 + n)\Delta f)t) \quad (2.55)$$

where $t_d \leq t \leq \tau + t_d$ and $0 \leq n \leq (N_{sc} - 1)$. A homodyne architecture is employed to simplify the expressions, but a heterodyne architecture may also be applied. The received baseband signal is

$$x_{BB}(t) = \cos \left(-2\pi(f_0 + (-(N_{sc} - 1)/2 + n)\Delta f)t_d + \pi \frac{\beta}{\tau} \left((t - t_d) - \frac{\tau}{2} \right)^2 \right) \quad (2.56)$$

where $t_d \leq t \leq \tau + t_d$, and $0 \leq n \leq (N_{sc} - 1)$. Assuming in-phase and quadrature channels, the received signal may be written in complex form:

$$x_{BB}(t) = \exp(-j2\pi(f_0 + (-(N_{sc} - 1)/2 + n)\Delta f)t_d) \exp \left(j\pi \frac{\beta}{\tau} \left((t - t_d) - \frac{\tau}{2} \right)^2 \right) \quad (2.57)$$

The target and radar are assumed stationary during the transmission and reception of the stepped chirp waveform. Berens [9] discusses the issue of relative motion and the compensation required to properly process the waveform.

Rearranging terms, the baseband signal in equation (2.57) is

$$x_{BB}(t) = \exp(-j2\pi f_0 t_d) \exp(-j2\pi p \Delta f t_d) \exp \left(j\pi \frac{\beta}{\tau} \left((t - t_d) - \frac{\tau}{2} \right)^2 \right) \quad (2.58)$$

where $p = n - (N_{sc} - 1)/2$, and $-(N_{sc} - 1)/2 \leq p \leq (N_{sc} - 1)/2$. With each pulse centered at baseband, the ADC has to support only the single-pulse instantaneous bandwidth or $F_s \geq \beta$. The reduced sampling rate is a key benefit of this waveform. As discussed

in Section 2.2.2, a lower sampling rate implies that the ADC is capable of supporting a higher ENOB. Sampling the continuous signal in equation (2.58) yields

$$x_{BB}(m) = \exp(-j2\pi f_0 t_d) \exp(-j2\pi p \Delta f_0 t_d) \exp\left(j\pi \frac{\beta}{\tau} \left(\left(\frac{m}{F_s} - t_d\right) - \frac{\tau}{2}\right)^2\right) \quad (2.59)$$

where $\lceil F_s t_d \rceil \leq m \leq \lceil F_s (\tau + t_d) \rceil$. The signal in equation (2.59) contains two phase terms, $2\pi f_0 t_d$ and $2\pi p \Delta f_0 t_d$, that are constant over a pulse. The second term, $2\pi p \Delta f_0 t_d$, varies pulse to pulse and must be accounted for when reconstructing the wider bandwidth signal within the signal processor.

The discussion has focused on a point target located at a particular time delay; however, a radar collects returns over a range window. To accumulate these returns, the receiver is active over the time interval

$$T_c = \tau + \frac{2\Delta R}{c} \quad (2.60)$$

The samples collected over the range window from the n -th pulse are $y_n(m)$ where

$$m = 0, \dots, \left\lceil F_s \left(\tau + \frac{2\Delta R}{c} \right) \right\rceil \quad (2.61)$$

and include the return collected from the point target in equation (2.59).

2.3.4 Processing Options

The objective is to create, within the signal processor, a composite waveform that exhibits a range resolution commensurate with composite waveform's transmit bandwidth by properly combining the N_{sc} baseband returns in equation (2.59). A means for interpolating in time and shifting in frequency the received pulses is required. In addition, careful consideration must be given to coherently combining the returns and accounting for phase differences between pulses. Both time-domain (TD) and frequency-domain (FD) approaches are described in the literature [7, 8, 10, 11] and are covered in subsequent sections. An example of a stepped chirp waveform processed in the frequency domain is also presented. The analysis focuses on the return from a point target, but the process is linear and is applicable to a superposition of returns at different delays.

2.3.4.1 Time Domain

The TD approach for processing a stepped chirp waveform is discussed in [7, 8], and the steps are outlined here for completeness:

1. When implementing the TD approach, it is assumed that the frequency step size is equal to the pulse bandwidth (i.e., $\Delta f = \beta$).
2. In the signal processor, the first step is to up-sample (i.e., interpolate) the baseband signals in equation (2.59) by a factor of N_{sc} . The interpolation may be implemented using a finite impulse response (FIR) filter. The interpolated signal is required to support the composite waveform bandwidth and is denoted $y'_n(m)$. The interpolated pulses have a frequency support defined by $-\pi N_{sc} F_s \leq \Omega \leq \pi N_{sc} F_s$.

3. A frequency shift is applied to each pulse by mixing the pulse with a complex sinusoid having a linear phase response. The mixer product is

$$z_n(m) = y'_n(m) \exp\left(j2\pi p \Delta f \frac{m}{N_{sc} F_s}\right) \quad m = 0, \dots, \left\lceil N_{sc} F_s \left(\tau + \frac{2\Delta R}{c}\right) \right\rceil \quad (2.62)$$

4. Next a phase correction is applied to each pulse to force a continuous phase transition between pulses. For an odd number of pulses, the phase correction applied to the n -th pulse is

$$\phi_n = 2\pi\beta\tau \sum_{i=1}^{((N_{sc}-1)/2)-n} i \quad \text{for } 0 \leq n \leq ((N_{sc}-1)/2) - 1 \quad (2.63)$$

and

$$\phi_n = 2\pi\beta\tau \sum_{i=1}^{n-(N_{sc}-1)/2} (i-1) \quad \text{for } ((N_{sc}-1)/2 + 1) \leq n \leq (N_{sc}-1) \quad (2.64)$$

To reduce the number of operations, the phase correction may be applied prior to the interpolation step.

5. The frequency-shifted and phase-corrected pulses are then time aligned via

$$z(m) = \sum_{n=0}^{N_{sc}-1} z_n(m - nN_{sc}F_s\tau) \quad (2.65)$$

The time alignment places a requirement on the product $F_s\tau$ to be an integer. Note that $z_n(m) = 0$ for $m < 0$ and for $m > [F_s(\tau + 2\Delta R/c)]$

2.3.4.2 Frequency Domain

The FD approach is defined in [10, 11]. In most instances, the frequency step size is chosen to be less than or equal to the pulse bandwidth (i.e., $\Delta f \leq \beta$). A continuous or overlapped frequency response prevents the introduction of gaps in the composite waveform spectrum, which would increase the range sidelobes.

The FD approach is introduced starting with the pulses received from a point target at time delay t_d . The pulses have been mixed to baseband and sampled. The spectrum of the sampled signal in equation (2.59) is

$$X_n(\omega) = \exp(-j2\pi f_0 t_d) \exp(-j2\pi p \Delta f t_d) X(\omega) \exp(-j\omega F_s t_d) \quad (2.66)$$

where $-\pi \leq \omega \leq \pi$ and

$$X(\omega) = DTFT \left\{ \exp \left(j\pi \frac{B}{\tau} \left(\frac{m}{F_s} - \frac{\tau}{2} \right)^2 \right) \right\} \quad 0 \leq m \leq [F_s\tau] \quad (2.67)$$

The filter matched to a pulse with spectrum $X(\omega)$ is $X^*(\omega)$, where the asterisk denotes the conjugate operator. Applying the filter to the pulses in equation (2.66)

$$Y_n(\omega) = \exp(-j2\pi f_0 t_d) \exp(-j2\pi p \Delta f t_d) |X(\omega)|^2 \exp(-j\omega F_s t_d) \quad (2.68)$$

Equation (2.68) represents the match filtered spectrum for the n -th pulse. On receive, each pulse is mixed and centered at baseband. To reconstruct the waveform, the pulses

are shifted by their respective frequency offsets, $p\Delta f$. Shifting the n -th spectrum by $2\pi p\Delta f/F_s$

$$Y_n \left(\omega - \frac{2\pi p\Delta f}{F_s} \right) = \exp(-j2\pi f_0 t_d) \exp(-j2\pi p\Delta f t_d) \left| X \left(\omega - \frac{2\pi p\Delta f}{F_s} \right) \right|^2 \exp(-j(\omega F_s - 2\pi p\Delta f) t_d) \quad (2.69)$$

and after reducing terms

$$Y_n \left(\omega - \frac{2\pi p\Delta f}{F_s} \right) = \exp(-j2\pi f_0 t_d) \left| X \left(\omega - \frac{2\pi p\Delta f}{F_s} \right) \right|^2 \exp(-j\omega F_s t_d) \quad (2.70)$$

where $(-\pi + 2\pi p\Delta f/F_s) \leq \omega \leq (\pi + 2\pi p\Delta f/F_s)$. Note that the phase term $2\pi p\Delta f t_d$ in equation (2.68) has been removed by the shifting operation. Its removal is necessary to coherently combine the pulses. The match filtered and frequency shifted spectra contain a constant phase term that is a function of f_0 and t_d and a linear phase term proportional to t_d . The linear phase term positions the response in time after the inverse transform is applied.

The spectra in equation (2.70) are a result of applying the DTFT. To compute the spectra inside a signal processor, the DFT is applied. The K -length DFT is constructed by sampling the spectrum in equation (2.70) at $\omega_k = 2\pi k/K$, $-K/2 \leq k \leq K/2 - 1$ to yield

$$Y_n \left(2\pi \frac{k}{K} - \frac{2\pi p\Delta f}{F_s} \right) = \exp(-j2\pi f_0 t_d) \left| X \left(2\pi \frac{k}{K} - \frac{2\pi p\Delta f}{F_s} \right) \right|^2 \exp \left(-j2\pi \frac{k}{K} F_s t_d \right) \quad (2.71)$$

To align the pulse spectra, the frequency step size must be an integer multiple of the DFT bin size. The size of a DFT bin is found by computing

$$\delta f_{DFT} = \frac{F_s}{K} \quad (2.72)$$

The frequency step size, Δf , is therefore constrained to be

$$\Delta f = P \delta f_{DFT} \quad (2.73)$$

where P is an integer. Substituting equations (2.72) and (2.73) into equation (2.71)

$$Y_n \left(2\pi \frac{k}{K} - 2\pi \frac{pP}{K} \right) = \exp(-j2\pi f_0 t_d) \left| X \left(2\pi \frac{k}{K} - 2\pi \frac{pP}{K} \right) \right|^2 \exp \left(-j2\pi \frac{k}{K} F_s t_d \right) \quad (2.74)$$

where $(-K/2 + pP) \leq k \leq (K/2 - 1 + pP)$. Shifting the spectra has increased the size of the DFT to $K + (N_{sc} - 1)P$, and the effective sample rate is $F_s \frac{K + (N_{sc} - 1)P}{K}$. Zero padding may be applied in the frequency domain to force a power of 2 size FFT or to further interpolate the match filtered time-domain response.

At this point, the spectra have been shifted, and a matched filter has been applied to each pulse. The next step is to properly stitch together the shifted spectra to form a composite spectrum that achieves the desired main lobe and sidelobe response when the inverse transform is applied. The stitching process is illustrated in Figure 2-9 using three

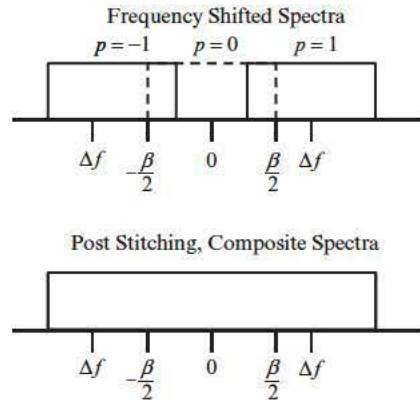


FIGURE 2-9 ■ The spectra of the stepped chirp pulses are stitched together to form a composite, rectangular-shaped spectrum.

pulses. A matched filter has been applied to each pulse, and the spectra are shifted by $-\Delta f$, 0 , and Δf , respectively. The trick is to stitch together the overlapping spectra in a manner that creates a “smooth” spectral response. An approach is to select the midpoint of the overlap between pulses n and $n + 1$ and to use the spectrum associated with pulse n for frequency bins to the left of (and including) the midpoint and the spectrum associated with pulse $n + 1$ to the right of the midpoint.

Having stitched together the subpulse spectra as illustrated in the lower plot of Figure 2-9, the shape of the composite spectrum approximates a rectangle. An inverse DFT is then applied to the spectrum to create the range compressed response. With a rectangular-shaped spectrum, the transform is a sinc. A weighting may be applied to the spectrum to reduce the range sidelobes. The composite bandwidth is

$$\beta_{sc} = \beta + (N_{sc} - 1) \Delta f \quad (2.75)$$

In the absence of a weighting, the Rayleigh range resolution achieved by the stepped chirp waveform is

$$\delta R = \frac{c}{2(\beta + (N_{sc} - 1) \Delta f)} \quad (2.76)$$

2.3.4.3 Summary of Frequency-Domain Processing Steps

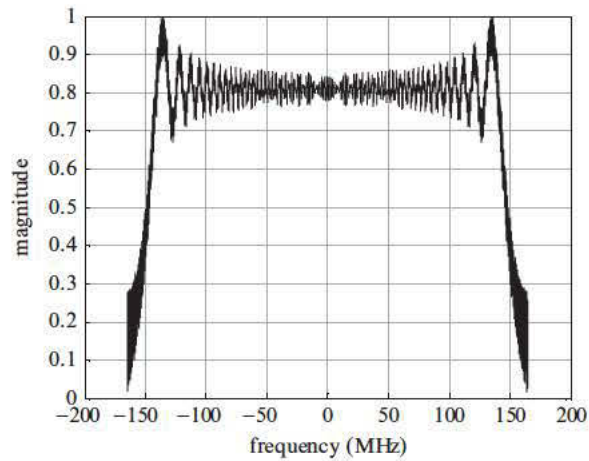
As noted in the previous section, several constraints are placed on the waveform when implementing FD processing:

1. The frequency step size must be an integer multiple of the DFT bin size to support proper alignment of the spectra.
2. The frequency step size is constrained such that $\Delta f \leq \beta$ to prevent gaps in the spectrum, which elevate the range sidelobes.

The following steps are implemented in the signal processor to reconstruct the waveform and to create a compressed response:

1. A DFT is applied to the received samples associated with each pulse.
2. A matched filter is applied to each pulse in the frequency domain.
3. The spectra are shifted by the appropriate frequency offset (as described in equation (2.74) for an odd number of pulses).
4. The shifted spectra are “stitched” together to create a composite spectrum approximating a rectangular shape.

FIGURE 2-10 ■ The stepped chirp waveform consists of three LFM pulses. The spectrum of a single LFM pulse with a 300 MHz bandwidth and a 1 μ sec pulse length is plotted.

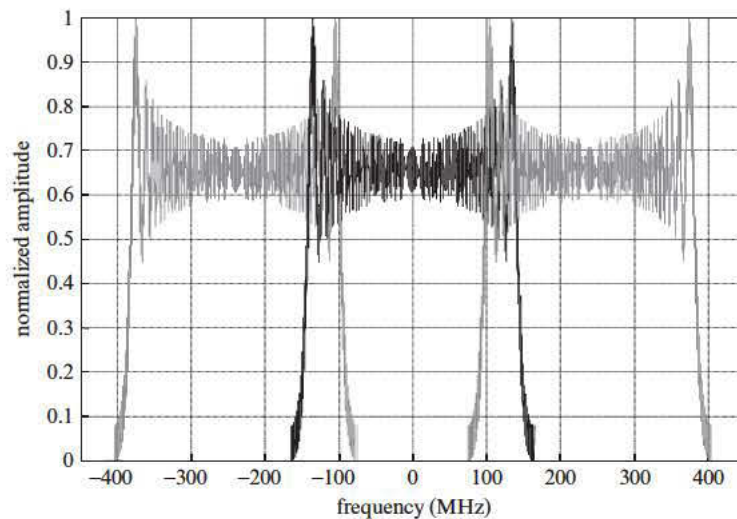


5. An amplitude weighting may be applied to reduce the range sidelobes.
6. Additional zeros may be appended to the spectrum to produce an interpolated response in the time domain and to satisfy a power of 2 or other size FFT requirement.
7. An inverse DFT is then applied to the composite spectrum to create the range compressed response.

2.3.4.4 Frequency-Domain Example

An example illustrates the resolution achieved by the waveform. Consider a stepped chirp waveform consisting of three LFM pulses each with a 300 MHz bandwidth and a 1 μ sec pulse length. The frequency step is 240 MHz. Figure 2-10 contains a plot of the spectrum associated with a single pulse centered at baseband. Each pulse exhibits the same spectrum except for a constant phase term as discussed in Section 2.3.4.2. A matched filter is applied to each waveform in the frequency domain, and the three spectra are shifted by the appropriate offset as shown in Figure 2-11. The spectra are then stitched together to

FIGURE 2-11 ■ The three spectra are aligned within the signal processor.



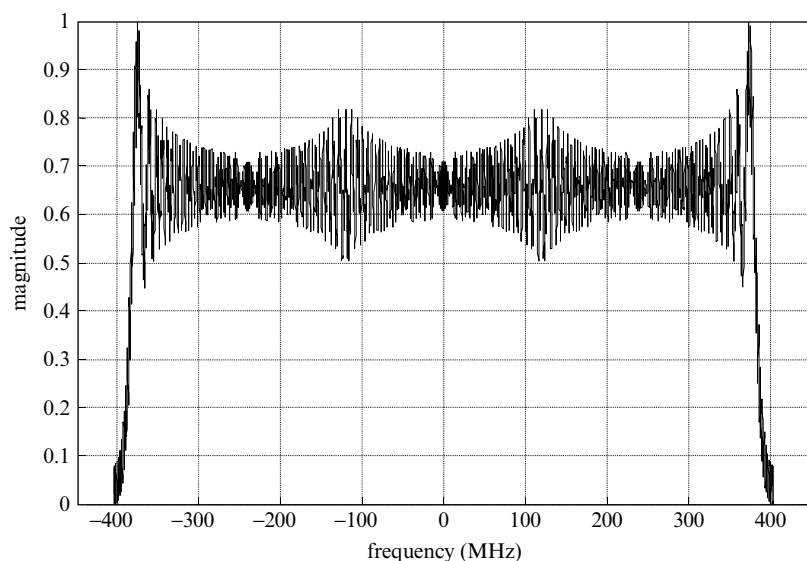


FIGURE 2-12 ■ The overlapping spectra are stitched together to form a composite spectrum.

create a smooth composite spectrum that resembles a rectangle. The stitching process is described in Section 2.3.4.2. Next, the range compressed response is obtained by taking the inverse DFT. A portion of the compressed response is shown in Figure 2-13, and the main lobe and adjacent sidelobes are plotted in Figure 2-14. Using the composite bandwidth defined in equation (2.75) and $\delta R = c/2\beta_{sc}$, the computed resolution is 0.192 meters, which is very close to the Rayleigh resolution achieved by the stepped chirp waveform in Figure 2-14. If desired, an amplitude taper may be applied to the composite spectrum in Figure 2-12 to reduce the range sidelobes.

2.3.5 Summary

The stepped chirp waveform is designed to reduce instantaneous bandwidth requirements on both transmit and receive. The waveform consists of several LFM pulses separated in frequency by a constant offset. Each pulse is mixed to baseband and sampled at a rate commensurate with the single-pulse bandwidth. The multiple-pulse waveform is reconstructed

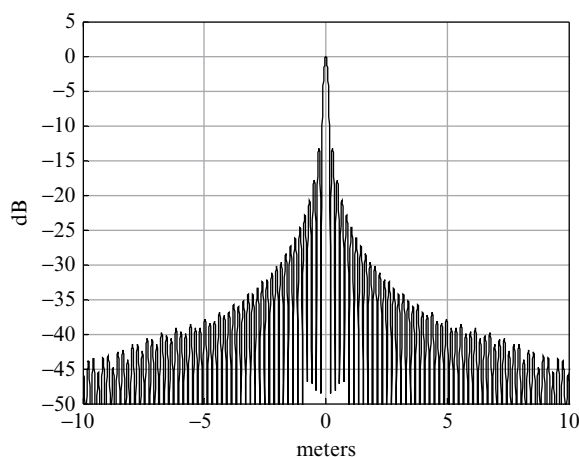
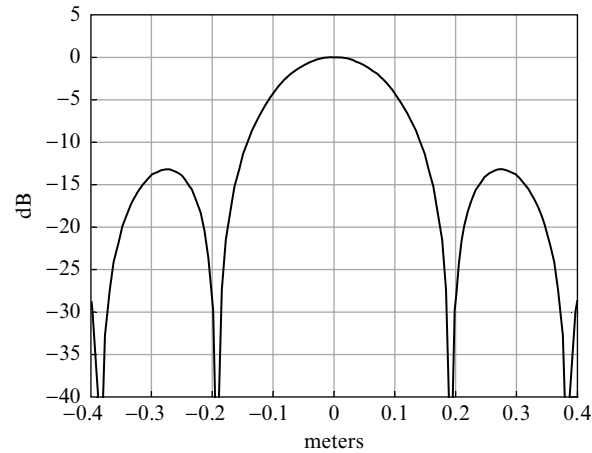


FIGURE 2-13 ■ The stepped chirp waveform's range compressed response resembles a sinc.

FIGURE 2-14 ■ The main lobe of the stepped chirp waveform exhibits enhanced range resolution.



within the signal processor to create a wider bandwidth signal with enhanced range resolution. The range window is not limited as in the case of stretch processing; however, a higher PRF may be required to support the additional pulses. In addition, compensation [9] for relative motion is required as part of the waveform reconstruction process.

2.4 | NONLINEAR FREQUENCY MODULATED WAVEFORMS

The relatively high range sidelobes associated with an LFM waveform are a by-product of the spectrum's shape, which approximates a rectangle. An amplitude taper may be applied to shape the spectrum and reduce the sidelobes at a cost of degraded resolution and a loss in SNR, the latter of which ranges from a few tenths of a dB for light weightings (e.g., -20 dB) to greater than 1 dB for heavier weightings (e.g., -40 dB). NLFM waveforms [19–26] achieve low range sidelobes by shaping the spectrum using frequency modulation. The modulation is applied over a constant time-domain envelope; therefore, a loss in SNR is not incurred. With both LFM and NLFM waveforms, degraded range resolution is associated with spectral shaping. NLFM waveforms exhibit range-Doppler coupling similar to that observed in an LFM waveform with the caveat that an NLFM waveform exhibits a degraded sidelobe response in the presence of a small fractional Doppler shift. In this text, the term *fractional Doppler* refers to the ratio of the Doppler shift to the waveform's intrapulse bandwidth. NLFM waveforms are considered less Doppler tolerant than their LFM counterparts.

The upper panel in Figure 2-15 contains a plot of an NLFM waveform's instantaneous frequency (solid line) as a function of time. The pulse width is τ , and the swept bandwidth is β . The bottom panel contains an idealized plot of an NLFM waveform's spectrum. The slope of the response is higher at the ends than in the middle of the pulse. The slope maps inversely into the amount of time spent at a given frequency. The more time spent at a particular frequency (corresponding to a shallow slope), the more power allocated to that frequency. An NLFM waveform shapes its spectrum by varying the slope of the instantaneous frequency. The tapered spectrum produces a match filtered response with range sidelobes defined by the shape of the square of the spectrum's envelope. In contrast, the slope of an LFM waveform's instantaneous frequency is a constant (dashed line in

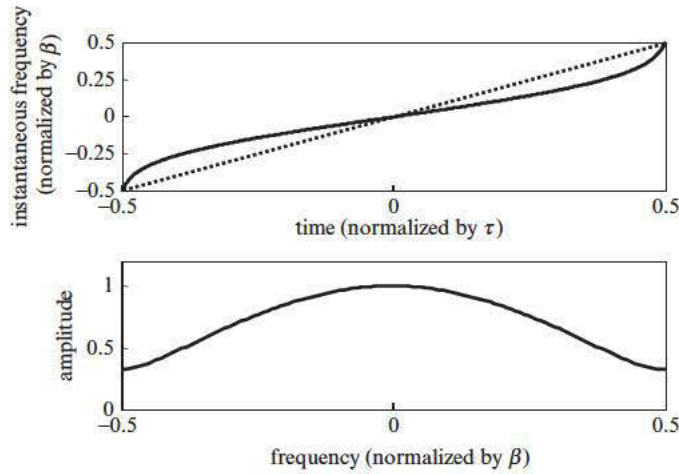


FIGURE 2-15 ■ The slope of the instantaneous frequency varies with time, producing a shaped spectrum. The upper box contains a plot of the instantaneous frequency for an NLFM (solid curve) and an LFM (dashed curve) waveform. The lower box contains a plot of a spectrum that might be generated by an NLFM waveform.

Figure 2-15). The LFM waveform distributes its power evenly over bandwidth, creating a rectangular shaped spectrum.

The design of an NLFM waveform consists of mapping the desired spectral shape into an instantaneous frequency response. Several approaches for designing and synthesizing NLFM waveforms are explored in the following sections.

2.4.1 Functional Relationships

Key and others [19, 20, 26] derived a set of relationships that aid in the design of NLFM waveforms. The process starts by defining a generic waveform with time-domain response

$$x(t) = a(t) \exp(j\phi(t)) \quad (2.77)$$

where $a(t)$ is the amplitude envelope, and $\phi(t)$ is the phase response. The waveform's spectrum is

$$X(\Omega) = |X(\Omega)| \exp(j\theta(\Omega)) \quad (2.78)$$

where $|X(\Omega)|$ is the spectrum's magnitude, and $\theta(\Omega)$ is the phase response. The time and spectral domains are related through the Fourier transform where

$$X(\Omega) = \int_{-\infty}^{\infty} a(t) \exp(j\phi(t)) \exp(-j\Omega t) dt \quad (2.79)$$

2.4.1.1 Principle of Stationary Phase

Key [19, 26] applies Kelvin's principle of stationary phase (PSP) to derive a set of parametric equations that relate the amplitude and phase functions in the two domains. The PSP states that the primary contribution to $X(\Omega)$ in equation (2.79) occurs when

$$\frac{d}{dt} (\phi(t) - \Omega t) = 0 \quad (2.80)$$

or

$$\Omega = \phi'(t) \quad (2.81)$$

where a single prime denotes the first derivative with respect to the independent variable. The relationship in equation (2.81) suggests that the derivative of the phase function (i.e., the instantaneous frequency) at a given value of time t determines the spectral response at a given frequency Ω and that Ω and t are one-to-one provided $\phi'(t)$ is monotonic.

Applying the inverse Fourier transform to the waveform spectrum

$$x(t) = \frac{1}{2\pi} \int |X(\Omega)| \exp(j\theta(\Omega)) \exp(j\Omega t) d\Omega \quad (2.82)$$

and again using the PSP, Key [19] derives a similar relationship

$$t = -\theta'(\Omega) \quad (2.83)$$

relating time and the derivative of the spectrum's phase.

Group delay is a measure of the relative time delay between a waveform's frequency components. Group delay, t_{gd} , is formally defined as the negative of the first derivative of the spectrum's phase function, or $t_{gd} = -\theta'(\Omega)$. If the group delay is a constant, then all frequencies are time coincident. If the group delay is linear, the frequency of the waveform varies linearly with time and produces a rectangular-shaped spectrum as in the case of an LFM waveform. Higher-order group delays may be used to shape a waveform's spectrum.

2.4.1.2 Inverse Functions

An inverse relationship between instantaneous frequency and group delay exists. To show this, solve equation (2.81) for t

$$t = \phi'^{-1}(\Omega) \quad (2.84)$$

Equating equations (2.83) and (2.84)

$$\phi'^{-1}(\Omega) = -\theta'(\Omega) \quad (2.85)$$

or equivalently

$$\phi'(t) = -\theta'^{-1}(t) \quad (2.86)$$

As illustrated in Figure 2-16, equations (2.85) and (2.86) define an inverse relationship between group delay and instantaneous frequency that is commonly exploited in the design of NLFM waveforms. Cook [26] states that the inverse relationship may be visualized by rotating a plot of the group delay (versus frequency) clockwise by 90° and flipping the result about the frequency axis to obtain a plot of instantaneous frequency versus time. The graphical technique may be applied in a computer to transform samples of a waveform's group delay versus frequency into an instantaneous frequency versus time response.

2.4.1.3 Parametric Equations

Key [19] uses equations (2.79), (2.81), (2.82), and (2.83), a Taylor series expansion, and several approximations to arrive at the following parametric relationships:

$$|X(\Omega_t)|^2 \approx 2\pi \frac{a^2(t)}{|\phi''(t)|} \quad (2.87)$$

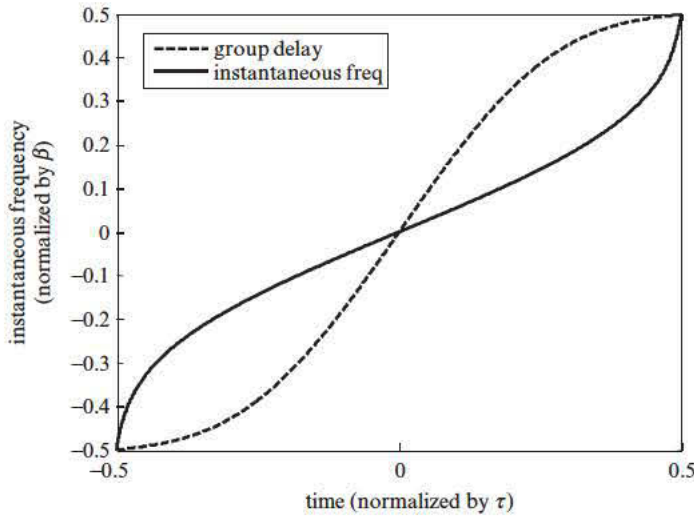


FIGURE 2-16 ■ An inverse relationship between instantaneous frequency and group delay exists, and the property may be exploited in the design of NLFM waveforms.

and

$$a^2(t_\Omega) \approx \frac{1}{2\pi} \frac{|X(\Omega)|^2}{|\theta''(\Omega)|} \quad (2.88)$$

where a double prime denotes the second derivative with respect to the independent variable. These equations provide a link between the waveform's time and frequency domains. The first derivative of the time-domain phase function, $\phi(t)$, is the waveform's instantaneous frequency, and in equation (2.87), the second derivative is the slope of the instantaneous frequency. For radar waveforms, the time-domain envelope, $a(t)$, is assumed to be constant. In equation (2.87), a relationship exists between the slope of the instantaneous frequency at a given time, t , and magnitude of the spectral response at a specific frequency, Ω_t . The subscript t affixed to Ω denotes the correspondence. Equation (2.88) defines a similar relationship between the second derivative of the frequency-domain phase response and the magnitude of the spectrum.

2.4.2 Design Approaches

To synthesize an NLFM waveform, a continuous or discrete instantiation of the time-domain phase function $\phi(t)$ is required. Several approaches [19–23, 26] for constructing $\phi(t)$ are described in the literature. Each of these approaches begins with the selection of an amplitude taper whose Fourier transform would produce the desired sidelobe response. Common tapers include Taylor and cosine on a pedestal weightings. Cosine on a pedestal tapers are of the form

$$W(\Omega) = h + (1 - h) \cos^2\left(\frac{\pi\Omega}{2\pi\beta}\right) \quad -\pi\beta \leq \Omega \leq \pi\beta \quad (2.89)$$

where the Hamming taper is defined for $h = 0.08$ and exhibits a peak sidelobe of -42.8 dB. The Taylor weighting function is defined in Section 2.4.3.

Given an amplitude taper and Key's [19] parametric equations, $\phi(t)$ may be synthesized using an inverse relationship between group delay and instantaneous frequency or using the inverse Fourier transform. Both approaches are covered in the following sections.

2.4.2.1 Inverse Function

Equations (2.85) and (2.86) define an inverse relationship between group delay and instantaneous frequency. Group delay is obtained by integrating a closed form expression of the amplitude taper. The relationship between $\theta''(\Omega)$ and $|X(\Omega)|$ is defined in equation (2.88). The resulting expression for group delay (i.e., $-\theta'(\Omega)$) is generally not invertible. However, Cook's [26] graphical inversion technique may be applied to samples of the group delay to construct a discrete time instantiation of $\phi'(t)$. The response may then be numerically integrated to obtain samples of $\phi(t)$. A closed-form expression may also be obtained by fitting the samples of $\phi'(t)$ to a model and integrating the response to obtain $\phi(t)$. An example is provided in Section 2.4.3.

2.4.2.2 Fourier Approach

Newton [21] recognized that the Fourier relationship in equation (2.82) could be used to construct $\phi(t)$ given $|X(\Omega)|$ and $\theta(\Omega)$. The spectrum's phase is obtained by integrating twice over the square of the spectrum magnitude. An inverse DFT is applied to samples of the spectrum to obtain discrete time samples of $a(t)$ and $\phi(t)$. In many instances, the transform produces a time-varying envelope $a(t)$, which Newton replaces with a constant. The spectrum of the time- and frequency-constrained waveform $\exp(j\phi(t))$ exhibits amplitude ripples that tend to elevate the range sidelobes. Newton applied a mismatched or inverse filter to compensate for the ripples. The loss associated with the mismatched filter (~ 0.2 dB) is significantly less than that associated with a traditional amplitude taper. Newton proposed combining an NLFM with a light taper to achieve the desired sidelobe response.

2.4.3 Design Example

A process for creating an NLFM waveform is demonstrated using the inverse relationship in equation (2.86) and the parametric relationship in equation (2.88). Consider a waveform with a constant time-domain envelope and a -40 dB peak sidelobe requirement. The waveform is expected to sweep its instantaneous frequency from $-\beta/2$ to $\beta/2$ during the time interval $-\tau/2 \leq t \leq \tau/2$ and $\beta\tau \gg 1$. The specific bandwidth and pulse length are undefined at the moment and are left as variables.

A Taylor weighting is chosen to meet the sidelobe requirement and is defined as

$$W_{Taylor}(\Omega) = G \left\{ 1 + 2 \sum_{m=1}^{\bar{n}-1} F_m \cos\left(\frac{m2\pi\Omega}{2\pi\beta}\right) \right\} \quad -\pi\beta \leq \Omega \leq \pi\beta \quad (2.90)$$

The coefficients F_m are

$$F_m = \begin{cases} \frac{(-1)^{m+1} \prod_{n=1}^{\bar{n}-1} \left[1 - \frac{m^2}{S(D^2 + (n-0.5)^2)} \right]}{2 \prod_{\substack{n=1 \\ n \neq m}}^{\bar{n}-1} \left(1 - \frac{m^2}{n^2} \right)}, & m = 1, 2, \dots, (\bar{n}-1) \\ 0, & m \geq \bar{n} \end{cases} \quad (2.91)$$

where

$$D = \frac{1}{\pi} \cosh^{-1} [10^{-PSR/20}] \quad (2.92)$$

$$S = \frac{\bar{n}^2}{D^2 + (\bar{n} - 0.5)^2} \quad (2.93)$$

and

$$G = \frac{1}{1 + 2 \sum_{m=1}^{\bar{n}-1} F_m} \quad (2.94)$$

The exact shape of the Taylor weighting is a function of the desired peak sidelobe ratio (PSR) and a parameter \bar{n} . The Taylor weighting in equation (2.90) is centered at baseband and extends over the frequency range $-\pi\beta \leq \Omega \leq \pi\beta$. The square of the spectrum magnitude in equation (2.88) is set equal to equation (2.90) or

$$|X(\Omega)|^2 = W_{Taylor}(\Omega) \quad (2.95)$$

The requirement for a constant time-domain envelope is satisfied by setting the envelope equal to 1 or

$$a^2(t) = 1 \quad (2.96)$$

Integrating equation (2.88) with respect to Ω yields

$$\theta'(\Omega) = \frac{k_1}{2\pi} G \left[\Omega + 2 \sum_{m=1}^{\bar{n}-1} \frac{\beta F_m}{m} \sin\left(\frac{m\Omega}{\beta}\right) \right] + k_2 \quad -\pi\beta \leq \Omega \leq \pi\beta \quad (2.97)$$

The waveform's group delay is

$$t_{gd} = -\theta'(\Omega) = -\frac{k_1}{2\pi} G \left[\Omega + 2 \sum_{m=1}^{\bar{n}-1} \frac{\beta F_m}{m} \sin\left(\frac{m\Omega}{\beta}\right) \right] - k_2 \quad -\pi\beta \leq \Omega \leq \pi\beta \quad (2.98)$$

The constants k_1 and k_2 are obtained by evaluating the group delay at the boundary conditions. Evaluating equation (2.98) at $t_{gd} = -\frac{\tau}{2}$ when $\Omega = -\pi\beta$, and at $t_{gd} = \frac{\tau}{2}$ when $\Omega = \pi\beta$, yields

$$k_2 = 0 \quad (2.99)$$

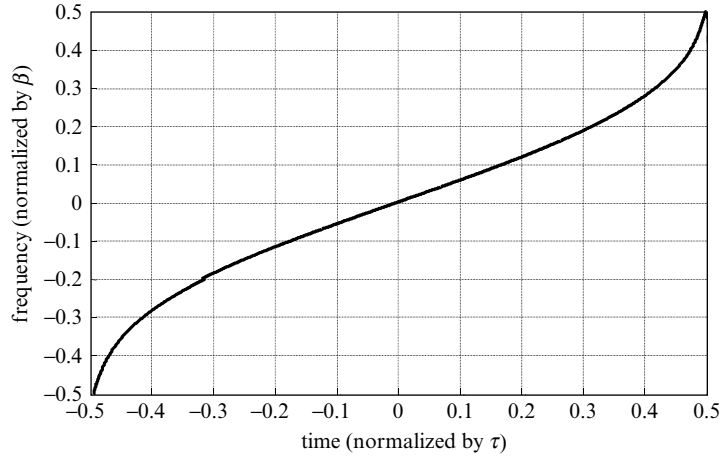
and

$$k_1 = -\frac{\tau}{G\beta} \quad (2.100)$$

Making the substitutions for k_1 and k_2 into equation (2.98),

$$t_{gd} = \frac{1}{2\pi} \frac{\tau}{\beta} \left[\Omega + 2 \sum_{m=1}^{\bar{n}-1} \frac{\beta F_m}{m} \sin\left(\frac{m\Omega}{\beta}\right) \right] \quad -\pi\beta \leq \Omega \leq \pi\beta \quad (2.101)$$

FIGURE 2-17 ■ The waveform's nonlinear frequency response will generate a Taylor weighted spectrum.



At this point, we want to implement the graphical technique proposed by Cook [26]. First, the group delay in equation (2.101) is evaluated at equally spaced points in Ω . Next, a numerical interpolation is performed on the samples to obtain equally spaced points in time (i.e., group delay) that map to specific values of the independent frequency variable, Ω . The time samples are then multiplied by -1 (flipping them about the group delay axis) and circularly rotated left to right (e.g., a clockwise rotation of a row vector containing the samples) creating new function with time as the new independent variable and instantaneous frequency as the new dependent variable. Figure 2-17 contains a plot of the instantaneous frequency obtained from equation (2.101) after applying a cubic interpolation and the inversion process.

A model for the instantaneous frequency [3] is proposed:

$$\phi'(t) = 2\pi\beta \left[\frac{t}{\tau} + 2 \sum_{m=1}^M d_k \sin \left(\frac{2\pi mt}{\tau} \right) \right] \quad (2.102)$$

Equation (2.102) asserts that the instantaneous frequency may be modeled as a linear term plus a sum of harmonically related and weighted sine functions. To obtain the coefficients d_k , the linear component is first subtracted from both sides of the equation

$$\frac{\phi'(t)}{2\pi\beta} - \frac{t}{\tau} = 2 \sum_{m=1}^M d_k \sin \left(\frac{2\pi mt}{\tau} \right) \quad (2.103)$$

The right side of equation (2.103) represents the first M terms of the Fourier series of an odd signal. Figure 2-18 contains a plot of the instantaneous frequency with the linear component removed. The response is odd and may be interpreted as one period of a periodic function. Harmonic analysis is applied to the signal in Figure 2-18 to obtain the coefficients d_k .

A periodic signal $x(t)$ may be expressed as a sum of weighted of sines and cosines where

$$x(t) = a_0 + 2 \sum_{m=1}^{\infty} [b_k \cos(m\Omega_0 t) + d_k \sin(m\Omega_0 t)] \quad (2.104)$$

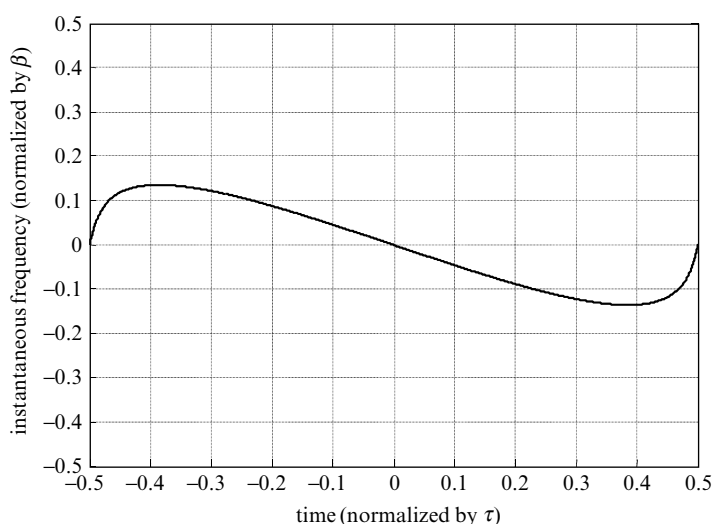


FIGURE 2-18 ■ The instantaneous frequency after removing the linear component is viewed as one period of a periodic function and is modeled using Fourier series coefficients.

where $\Omega_0 = 2\pi/\tau$. For odd signals

$$x(t) = 2 \sum_{m=1}^{\infty} d_k \sin(m\Omega_0 t) \quad (2.105)$$

The Fourier series coefficients d_k are obtained by evaluating the integral

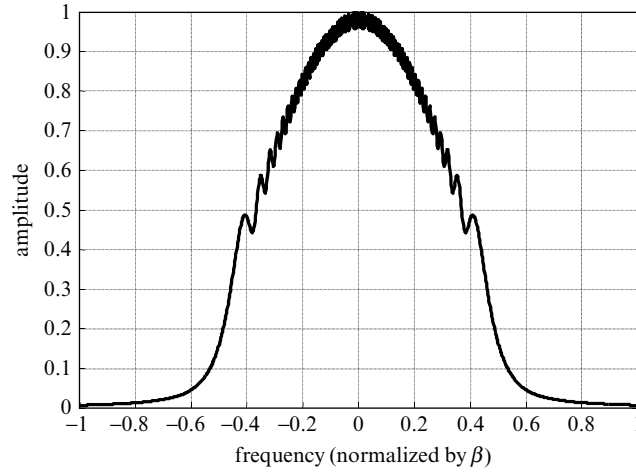
$$d_k = \int_{\tau} x(t) \sin(m\Omega_0 t) dt \quad (2.106)$$

where τ is one period of the waveform or, in this case, the pulse width. The coefficients are obtained numerically by replacing the integral in (2.106) with a summation. Table 2-2

TABLE 2-2 ■ Coefficients Used to Generate an NLFM Waveform

Index	Coefficient Value	Index	Coefficient Value
1	-0.0570894319	16	0.0003386501
2	0.0198005915	17	-0.0002874351
3	-0.0102418371	18	0.0002451937
4	0.0062655130	19	-0.0002100998
5	-0.0042068327	20	0.0001807551
6	0.0029960800	21	-0.0001560755
7	-0.0022222723	22	0.0001352110
8	0.0016980153	23	-0.0001174887
9	-0.0013271886	24	0.0001023710
10	0.0010560642	25	-0.0000894246
11	-0.0008525769	26	0.0000782981
12	0.0006965787	27	-0.0000687043
13	-0.0005748721	28	0.0000604069
14	0.0004785116	29	-0.0000532106
15	-0.0004012574	30	0.0000469533

FIGURE 2-19 ■ The NLFM waveform exhibits a shaped spectrum resembling a Taylor weighting. The amplitude ripple is associated with the finite time and frequency extent constraints placed on the waveform.



contains the first 30 coefficients. No attempt is made to determine the number of coefficients or the numerical precision needed to achieve a specified level of performance.

The time-domain phase function is obtained by integrating the expression in equation (2.102)

$$\phi(t) = \frac{\pi\beta}{\tau}t^2 - \sum_{m=1}^M \frac{\beta\tau 2d_k}{m} \cos\left(\frac{2\pi mt}{\tau}\right) \quad (2.107)$$

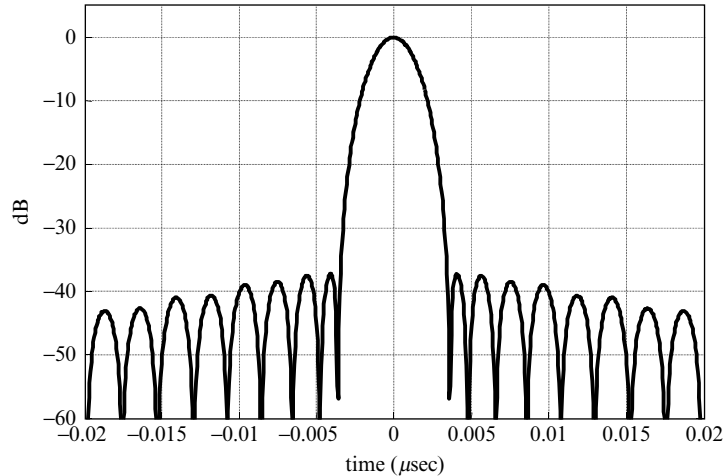
The resultant baseband, complex NLFM waveform is

$$x(t) = \exp(j\phi(t)) \quad (2.108)$$

The waveform's spectrum is plotted in Figure 2-19. The envelope of squared spectrum resembles a Taylor weighting. The ripples in the spectrum are a result of the waveform's finite extent imposed in both domains.

Consider an LFM waveform with a 500 MHz swept bandwidth and a 1 μ sec pulse length. Applying the matched filter in either the time or frequency domain generates the compressed response shown in Figure 2-20. The peak sidelobes are approximately

FIGURE 2-20 ■ The NLFM waveform achieves a range compressed response with peak sidelobes slightly above -40 dB.



−37 dB; the goal was −40 dB. The peak sidelobe levels are close to the designed-to level. The waveform's time–bandwidth product does have a direct bearing on sidelobe performance. In general, the higher the time–bandwidth product the closer one gets to the design-to level. Time–bandwidth products greater than 100 are generally required.

For an LFM waveform with 500 MHz swept bandwidth, the time resolution, measured at the −4 dB width, is 0.002 μsec . A close examination of the response in Figure 2-20 reveals a −4 dB width of 0.0029 μsec . The modulation has degraded the resolution by a factor of 1.45 over that achieved by an LFM swept over the same bandwidth. Of course, a similar degradation in resolution occurs when an amplitude taper is applied to an LFM waveform.

2.4.4 Doppler Analysis

NLFM waveforms are less Doppler tolerant [24–26] than their LFM counterparts. NLFM waveforms exhibit high near-in sidelobes in the presence of small, uncompensated, fractional Doppler shifts (i.e., f_d/β). Doppler shifts on the order of a 1/4 cycle across the uncompressed pulse width are sufficient to significantly elevate the near-in sidelobes. Figure 2-21 and 22 contain plots of the compressed response with a 1/4 and full cycle of Doppler, respectively. The Doppler shifts used in this example (0.25 MHz and 1 MHz, respectively) are not commonly realized, but several cycles of Doppler may be observed when employing longer pulses (e.g., 1 msec) operating against fast moving targets. In this example, the peak sidelobes have increased from −40 dB to approximately −28 dB and −15 dB, respectively. As with the LFM waveform, a loss in the peak amplitude and resolution are experienced with increasing Doppler shift. Johnston [24, 25] provides an excellent discussion on Doppler sensitivity.

2.4.5 Summary

NFLM waveforms achieve low range sidelobes without the SNR loss associated with an amplitude taper. Frequency modulation is used to shape the waveform's spectrum, and the square of the spectrum magnitude determines the shape of the time-domain response.

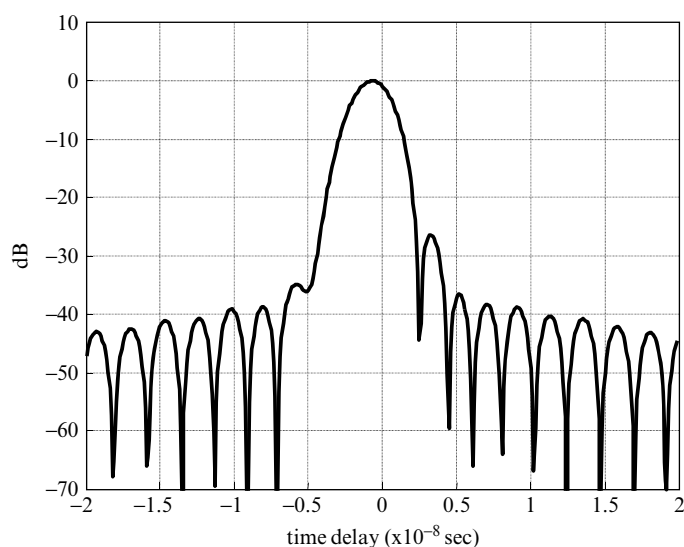
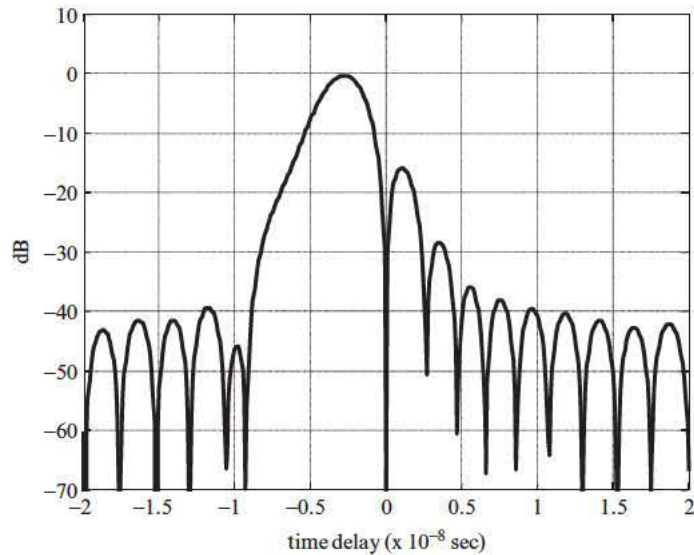


FIGURE 2-21 ■ A quarter cycle of Doppler shift across the uncompressed pulse starts to elevate the near-in sidelobes of the NLFM waveform.

FIGURE 2-22 ■ A full cycle of Doppler shift across the uncompressed pulse increases the near-in sidelobes of the NLFM waveform to a peak value of -15 dB and displaces the main lobe by one range resolution bin.



NLFM waveforms are less Doppler tolerant than LFM waveforms, and stretch processing is not an option for NLFM waveforms. The latter two properties may on occasion limit its application.

2.5 | STEPPED FREQUENCY WAVEFORMS

In some applications, a waveform with the following properties is advantageous:

1. A low instantaneous bandwidth on transmit and receive
2. A composite waveform bandwidth that is greater than the waveform's instantaneous (intrapulse) bandwidth and that defines the range resolution achieved by the system
3. A relatively short blind range

Waveforms with these properties are beneficial in radar upgrades where additional bandwidth is desired but where instantaneous bandwidth is limited, in phase-steered antennas where instantaneous bandwidth increases dispersion loss and other frequency dependent effects, and in wideband instrumentation radars where the target/clutter environment is generally more constrained and known over relatively long time intervals.

An SF waveform [16–18] achieves the aforementioned properties by transmitting and receiving N narrowband pulses each at a different center frequency and coherently combining the returns to create a waveform whose composite bandwidth is greater than the single-pulse bandwidth. Since the waveform's bandwidth and energy are spread over several pulses, the duration of a single pulse and resultant blind range are reduced. The advantages associated with an SF waveform are similar those ascribed to a stepped chirp waveform in Section 3, but the processing steps and some of the fundamental properties are very different. Disadvantages associated with an SF waveform include the additional time required to transmit and receive N pulses and range ambiguities introduced as a by-product of the frequency stepping. These shortcomings tend to constrain its application; however, there are instances where the waveform is the correct one to employ.

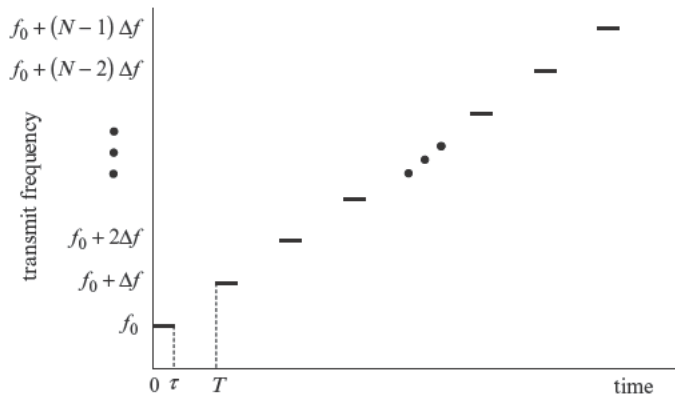


FIGURE 2-23 ■ A stepped frequency waveform consists of multiple pulses separated in frequency by a step size Δf .

2.5.1 Waveform Description

An SF waveform is composed of N pulses separated from one another in frequency by Δf hertz and in time by T seconds as illustrated in Figure 2-23. The radar transmits a pulse and then collects samples at specific range gates (or time delays). The process is repeated for N pulses. A high-resolution range profile is then formed at each range gate. After processing, the resultant range resolution is inversely proportional to the composite waveform bandwidth, $N\Delta f$.

With an SF waveform, there is no restriction on the intrapulse modulation; a continuous wave (CW) pulse or a pulse employing phase or frequency modulation may be employed. However, as a part of the trade space, waveform parameters such as pulse width, bandwidth, PRI, and frequency step size must be selected based on system requirements and the anticipated target and clutter environment.

2.5.2 Phase Rotation Used to Measure Range

By transmitting pulses centered at different frequencies, an SF waveform exploits phase changes to resolve returns in range at resolutions finer than that achieved with a single pulse. For a stationary point target located at range R_0 , the measured phase on the first pulse is

$$\theta = 2\pi f_0 \frac{2R_0}{c} \quad (2.109)$$

where f_0 is the transmit frequency, and c is the speed of light. Stepping the center frequency from pulse to pulse by an amount Δf produces a phase rotation

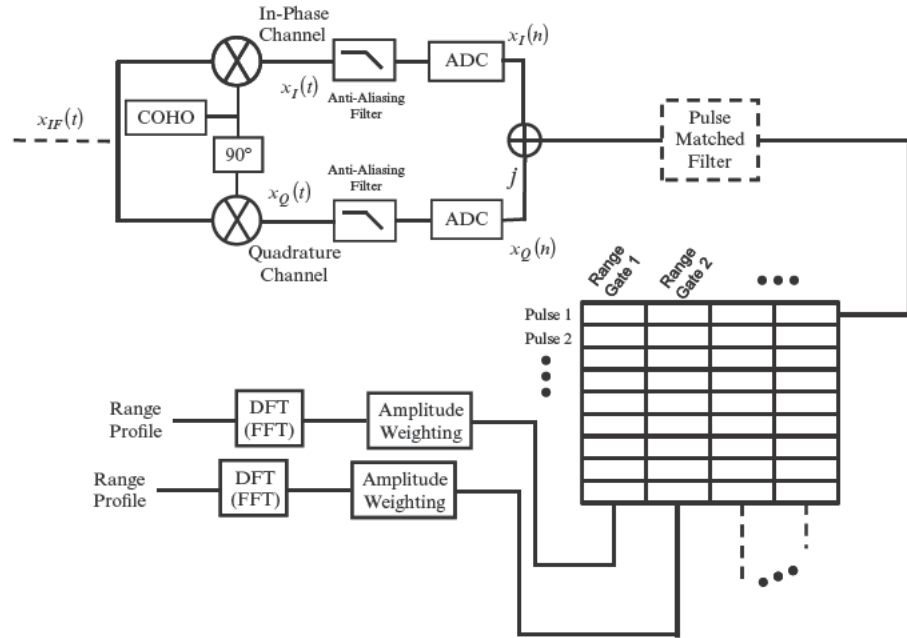
$$\theta(n) = 2\pi f_0 \frac{2R_0}{c} + 2\pi n \Delta f \frac{2R_0}{c} \quad n = 0, \dots, N-1 \quad (2.110)$$

where n is the pulse index. The phase rotates at a rate

$$\Delta\theta = \theta(n+1) - \theta(n) = 2\pi \Delta f \frac{2R_0}{c} \quad (2.111)$$

and provides a mechanism for resolving targets in range. As discussed in Section 2.5.4.2, the returns are processed using a DFT, which maps the phase rotations into an HRR profile.

FIGURE 2-24 ■ An HRR profile is formed at each range gate using a DFT to compress the returns from N pulses.



2.5.3 Receiver and Processing Architecture

The receiver, the fast and slow time samples, and the signal processing blocks associated with an SF waveform are depicted in Figure 2-24. In the receiver, the returns from the n -th pulse are mixed from RF to an intermediate frequency and finally to baseband using a coherent detector. A matched filter is applied to the received pulse either in the form of a low-pass filter preceding the ADC for a CW pulse or as a matched filter implemented in the signal processor for a modulated pulse.

As returns from a pulse are received, a sample is collected at each range gate, and the process is repeated over N pulses. The columns forming the matrix in Figure 2-24 represent M range gates collected over fast time, and the rows index returns from the n -th pulse over slow time. After collecting returns from N pulses, a range profile is formed at each range gate by processing the returns using a DFT. Range sidelobes are reduced by applying an amplitude weighting.

A single range profile contains resolved returns associated with scatterers located within a region centered on the range gate and extending a distance defined by twice the range resolution associated with a single pulse. Multiple range profiles may be concatenated to form a composite range profile representing the returns over several range gates.

2.5.4 Creating a Range Profile

2.5.4.1 Compressed Range Response

The range compression or enhanced resolution associated with an SF waveform is achieved by coherently combining the returns collected from N pulses. For a point target located at range R_0 , the samples are

$$x(n) = \exp\left(j2\pi f_0 \frac{2R_0}{c}\right) \exp\left(j2\pi n \Delta f \frac{2R_0}{c}\right) \quad n = 0, \dots, N-1 \quad (2.112)$$

The first term to the right of the equal sign is a constant and is associated with the round-trip delay and the initial transmit frequency. The second term is a complex phasor rotating at a rate defined by the frequency step size and the range to the target. The phase rotation is similar to that observed when employing a pulsed Doppler waveform. To simplify the expressions, the amplitude of the return is set to unity.

Fourier analysis is applied to the complex samples in equation (2.112) to extract the location of the return in range. Consider the DTFT defined by

$$X(\omega) = \sum_{n=0}^{N-1} x(n) \exp(-j\omega n) \quad (2.113)$$

where $x(n)$ are the measured returns collected from N pulses, and ω is the digital frequency with units of radians/sample. The DTFT represents a filter bank tuned over a continuum of frequencies (or rotation rates), which in this case correspond to different ranges.

The samples are often viewed as being collected in the frequency domain, and the returns are then transformed to the time (or range) domain. The frequency-domain interpretation is based on the assertion that each pulse is measuring the target's response (amplitude and phase) at a different frequency. From this perspective, an inverse DTFT would naturally be applied; however, it is the rotating phase induced by the change in frequency that is important. Either a forward or inverse DTFT may be applied as long as the location of the scatterers (either up- or down-range) is correctly interpreted within the profile, and the return is scaled to account for the DTFT integration gain.

For a point target located at range R_0 , the output of the DTFT is a digital sinc defined by

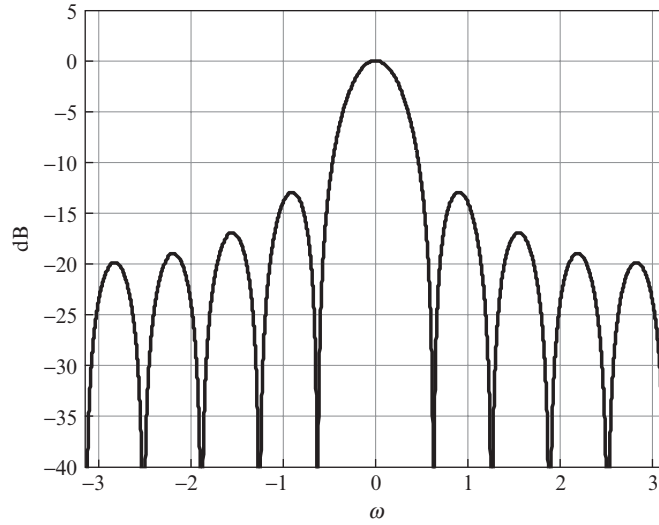
$$|X(\omega)| = \left| \frac{\sin\left(\frac{N}{2}(\omega - \omega_{R_0})\right)}{\sin\left(\frac{(\omega - \omega_{R_0})}{2}\right)} \right| \quad (2.114)$$

where $\omega_{R_0} = 2\pi \frac{2R_0}{c} \Delta f$, $\omega = 2\pi \frac{2R}{c} \Delta f$, and R is a continuous variable representing range. Equation (2.114) represents the range compressed response. The term ω_{R_0} centers the response at a particular range or frequency. The shape of the compressed response is examined by setting $R_0 = 0$ or

$$|X(\omega)| = \left| \frac{\sin\left(\frac{N}{2}\omega\right)}{\sin\left(\frac{\omega}{2}\right)} \right| \quad (2.115)$$

A plot of the compressed response is shown in Figure 2-25. The response consists of a main lobe and sidelobe structure with peak sidelobes 13.2 dB below the peak of the main lobe. The DTFT is periodic and repeats at multiples of 2π in ω ; therefore, the range compressed response is also periodic with periodicities spaced by $c/2\Delta f$. An implication is that range measured at the output of the DTFT is relative to the range gate and not absolute range. Absolute range is defined by the time delay to the range gate and the relative range offset within the profile.

FIGURE 2-25 ■ A SF waveform's compressed response is a digital sinc function.



The waveform's Rayleigh resolution is found by setting the argument of the numerator in equation (2.115) equal to π and solving for ω

$$\delta\omega = \frac{2\pi}{N} \quad (2.116)$$

where $\delta\omega$ is the resolution in terms of digital frequency. To convert the frequency resolution in equation (2.116) to a range resolution, consider two point targets located at ranges R_1 and R_2 and separated in range by $\delta R = |R_2 - R_1|$. The difference in their phase rotation rates is $2\pi \frac{2\delta R}{c} \Delta f$. Equating the rate difference to the frequency resolution defined in equation (2.116)

$$2\pi \frac{2\delta R}{c} \Delta f = \frac{2\pi}{N} \quad (2.117)$$

and solving for the range difference yields the Rayleigh resolution

$$\delta R = \frac{c}{2N\Delta f} \quad (2.118)$$

The range resolution achieved by the SF waveform is inversely proportional to the waveform's composite bandwidth $N\Delta f$. In this case, the Rayleigh resolution is equivalent to the main lobe's -4 dB width.

2.5.4.2 Discrete Fourier Transform

The DTFT is defined over continuous frequency; however, to realize the compressed response in digital hardware requires one to evaluate the DTFT at a finite number of frequencies over the interval $[0, 2\pi)$. It is common to evaluate the DTFT at equally spaced frequencies

$$\omega_k = 2\pi \frac{k}{M} \quad k = 0, \dots, (M-1) \quad (2.119)$$

where $M \geq N$. This is equivalent to evaluating the response at equally spaced ranges defined by

$$R_k = \frac{c}{2\Delta f M} k \quad k = 0, \dots, (M - 1) \quad (2.120)$$

Inserting equation (2.119) into (2.113) yields the DFT

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(-j \frac{2\pi k}{M} n\right) \quad M \geq N \quad (2.121)$$

which is often implemented using an FFT for computational efficiency. For a point target located at range R_0 , the compressed response is a sampled instantiation of the digital sinc in (2.114) or

$$|X(k)| = \left| \frac{\sin\left(\frac{N}{2} \left(\frac{2\pi}{M} k - \omega_{R_0}\right)\right)}{\sin\left(\left(\frac{2\pi}{M} k - \omega_{R_0}\right) / 2\right)} \right| \quad k = 0, \dots, (M - 1) \quad (2.122)$$

The DFT is a linear operator; thus, the range profile associated with multiple scatterers is simply the superposition of the individual responses.

Zero padding is often used in conjunction with a DFT to decrease the filter spacing, which reduces straddle loss. For $M > N$, the sequence is said to be zero padded with $M - N$ zeros. For $M = N$, the filters or range bins are spaced by the nominal waveform resolution $c/2N\Delta f$.

2.5.4.3 Range Sidelobe Suppression

Range sidelobes associated with large RCS targets may mask the presence of smaller targets. As noted in Figure 2-24, sidelobes are suppressed by applying an amplitude weighting to the complex samples collected at a range gate. When selecting an amplitude taper, the associated reduction in resolution and loss in SNR should be taken into account [36].

2.5.5 Straddle Loss and Range Ambiguities

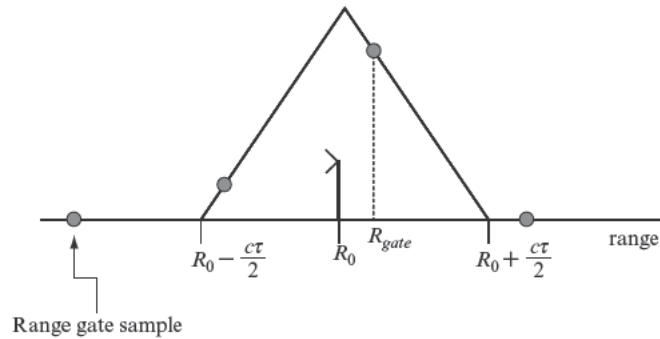
The range profile generated by an SF waveform may exhibit undesired artifacts as a result of ambiguities and straddle loss. The following sections examine these issues and the impact of waveform parameter selection on performance.

2.5.5.1 Straddle Loss

Straddle loss is introduced at two points in the processing chain: (1) when sampling the pulse's match filtered response; and (2) within the DFT. The loss associated with the DFT was discussed in Section 2.5.4.2. The loss associated with sampling is examined in this section. When sampling the return from a pulse, the resultant straddle loss is a function of the range gate spacing and the single-pulse bandwidth or range resolution. Several examples are provided to illustrate the loss.

In a radar system, range gates are often spaced by the transmit pulse's range resolution. A similar spacing may be applied to a stepped frequency waveform where the range gates

FIGURE 2-26 ■ A point target's matched filtered response is sampled at the range gate spacing.



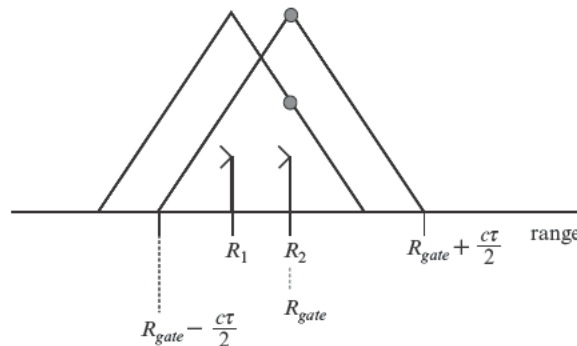
are separated by $c\tau/2$ when employing a CW pulse. Consider the match filtered response for an ideal CW pulse as shown in Figure 2-26. The response extends over a time delay equal to twice the pulse length (i.e., 2τ). For a point target, the peak of the response is located at the target's range R_0 . Four range gates, spaced by $c\tau/2$, are depicted in Figure 2-26. The darkened circles represent the samples collected by the radar and are aligned with the range gates. A range gate is not constrained to fall on the peak of the response, and the offset in range between the target, R_0 , and the range gate, R_{gate} , introduces straddle loss.

When forming an HRR profile of a target, straddle loss reduces the apparent RCS of individual scatterers resulting in a distorted response. To appreciate the impact, let's examine two cases: (1) a target consisting of only two scatterers; and (2) a target of length L consisting of multiple scatterers.

Consider two equal amplitude scatterers located at ranges R_1 and R_2 , as shown in Figure 2-27. The radar samples the composite response at range R_{gate} . The second scatterer, centered at range R_2 , is aligned with the range gate and is sampled at the peak of its response. In contrast, the first scatterer experiences a loss due to its misalignment with the range gate. The amount of loss is a function of the scatterer's distance from the range gate and the shape of the pulse's response.

Next, consider a target of length L , as shown in Figure 2-28. The responses for three scatterers, located in the center (solid line) and both ends of the target (dashed lines), are depicted. Since the range gate and the center scatterer are aligned, no straddle loss occurs. If the pulse width is chosen such that $c\tau/2 = L$, the scatterers at the ends of the target experience a 6 dB loss in SNR or measured RCS. When transmitting a phase code with rectangular chips, the loss is identical to that of a CW pulse. When employing an LFM pulse with bandwidth β and $c/2\beta = L$, the loss is ≤ 4 dB depending on the weighting applied.

FIGURE 2-27 ■ Scatterers located on a range gate are sampled at the peak of the response. Scatterers not located on a range gate experience straddle loss.



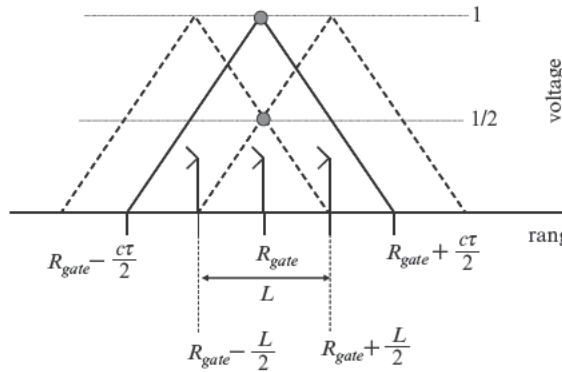


FIGURE 2-28 ■ An extended target will experience straddle loss that varies along the length of the target.

For an isolated target, the loss may be minimized over the target extent using a pulse with bandwidth such that $c/2\beta > L$. In a target-rich or extended-clutter environment, the pulse bandwidth must be carefully chosen to minimize the amount of energy associated with scatterers located in adjacent ambiguities that folds into the region of interest and produces ambiguous returns. Range ambiguities are discussed in the next section.

2.5.5.2 Range Ambiguities Associated with Frequency Stepping

An SF waveform exhibits range ambiguities that are a by-product of the frequency stepping. These ambiguities should not be confused with those associated with the PRI, which are also present. The ambiguity presents itself in equation (2.112), which is periodic in 2π . The periodicity causes returns located at $R_0 + mR_A$ where $m = \dots -2, -1, 0, 1, 2, \dots$ to rotate in phase at the same rate as the target located at range R_0 . The ambiguous range offset R_A is found by setting the phase rotation equal to 2π or

$$2\pi \Delta f \frac{2R_A}{c} = 2\pi \quad (2.123)$$

and solving for

$$R_A = \frac{c}{2\Delta f} \quad (2.124)$$

The phase rotation rate repeats or is ambiguous every $c/2\Delta f$ in range.

As an example, consider the response associated with two point targets located at ranges R_0 and $R_0 + c/2\Delta f$, as shown in Figure 2-29. In this example, the pulse width

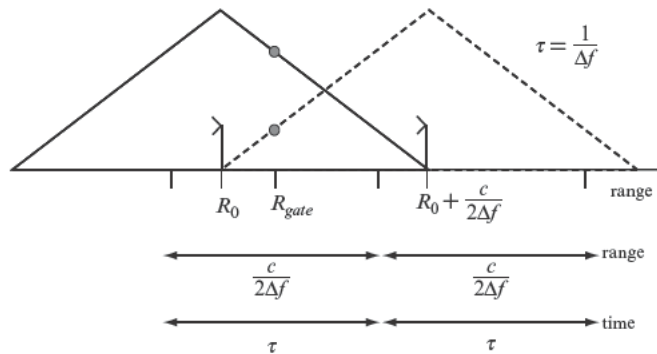
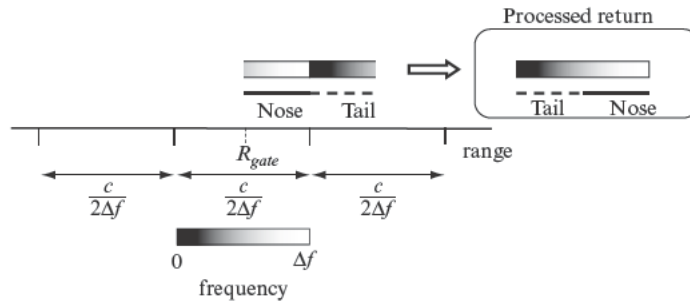


FIGURE 2-29 ■ A stepped frequency waveform introduces ambiguities in range that are separated by $c/2\Delta f$.

FIGURE 2-30 ■ The size of the unambiguous region must be larger than the target extent to prevent unresolvable ambiguities.



is $1/\Delta f$. The range gate is centered in the first unambiguous region and is labeled R_{gate} . The match filtered responses overlap even though the two targets lie in separate regions. Since the targets are separated by $c/2\Delta f$, their rotation rates are equivalent. The resultant range profile will contain a compressed response at range R_0 with an amplitude and phase that is equal to the coherent sum of the two responses.

To increase the size of the unambiguous region, the size of Δf may be reduced; however, for a given bandwidth, reducing the frequency step size increases the number of pulses and thus the time associated with the waveform. In a dynamic environment, a limited dwell time is required to prevent returns from decorrelating. For example, a target moving through one or more resolution bins during a collection interval causes the response to broaden in range and results in a reduction in achieved SNR. Trades associated with Δf and τ are examined in the following sections.

2.5.5.3 Ambiguities in a Dynamic Target Environment

In a dynamic environment, the location of the target relative to a range ambiguity is not constrained. In many cases, the target may straddle two unambiguous regions, as illustrated in Figure 2-30. The solid and dashed horizontal line segments represent scatterers positioned along an extended target. Assume that the two line segments represent an aircraft with its nose oriented toward the left side of the page. The nose section is represented by the solid line, and the tail section is represented by the dashed line. The unambiguous regions are spaced by $c/2\Delta f$. The phase rotation rate at the beginning of an unambiguous region is associated with 0 hertz, and the phase rotation rate at the upper end of an unambiguous region is slightly less than Δf hertz.

The pulse is sufficiently long to cover the target extent and extends over two ambiguities to minimize straddle loss. The shaded box illustrates how range and frequency are coupled over an unambiguous region. The darker and lighter regions correspond to lower and higher frequencies, respectively. Given the relative position of the target, the nose of the aircraft is at a higher frequency than the tail section. Having collected N samples and compressed the response, the processed return has the nose and tail sections reversed in the range profile. To resolve the ambiguity, it is necessary to select $c/2\Delta f > L$.

2.5.5.4 Choosing a Pulse Width in a Target- or Clutter-Rich Environment

In a target- or clutter-rich environment, it is desirable to limit the returns in a range gate to those associated with objects contained within a region extending $\pm c/4\Delta f$ about the center of the gate. Without this constraint, returns from neighboring ambiguous regions may fold into the range gate of interest. For a CW pulse, this condition is met by equating

the size of the unambiguous region to the post-match filtered pulse length 2τ or

$$2\tau = \frac{1}{\Delta f} \quad (2.125)$$

and solving for τ yields

$$\tau = \frac{1}{2\Delta f} \quad (2.126)$$

With a phase or frequency modulated pulse, one option is to set the null-to-null range resolution equal to the unambiguous range extent. For an unweighted LFM pulse, this is equivalent to

$$2\frac{c}{2\beta} = \frac{c}{2\Delta f} \quad (2.127)$$

or

$$\beta = 2\Delta f \quad (2.128)$$

With modulated pulses, low range sidelobes are used to suppress returns from objects located outside the unambiguous region.

Using the four scatterers in Figure 2-31, the consequences associated with selecting $\tau = 1/2\Delta f$ are examined. Scatterer #1 is positioned on a range gate. Scatterer #2 is located a distance of $c/8\Delta f$ from the first scatterer, and similarly scatterer #3 is placed a distance of $c/4\Delta f$ away. Scatterer #4 is situated at the center of the next range gate, which is spaced $c/2\Delta f$ from the first gate. With $\tau = 1/2\Delta f$, measurements taken at a range gate are associated only with scatterers located within a distance of $\pm c/4\Delta f$ about the gate. The pulse width prevents ambiguous returns from adjacent regions from folding into the measurement. A drawback is that scatterers located a distance $c/8\Delta f$ from a range gate experience a straddle loss of 6 dB. For scatterers located between $R_{gate} + c/8\Delta f$ and $R_{gate} + c/4\Delta f$, straddle loss is even more significant. The waveform designer is presented with a trade-off between suppressing ambiguous returns and reducing straddle loss. How the designer addresses the trade is dependent on the target and clutter environment. To

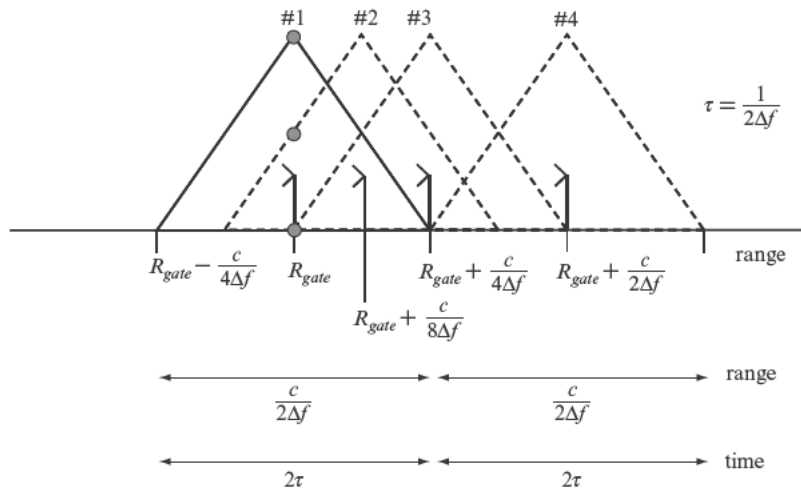


FIGURE 2-31 ■ Reducing the pulse width aids in dealing with range ambiguities but increases straddle loss.

address straddle loss, additional range gates may be inserted within an unambiguous region, but the result is a higher sampling rate and increased processing.

2.5.5.5 Waveform Parameter Trade Summary

An SF waveform is defined by four principal parameters, whose impact on performance is interrelated: pulse repetition interval, number of pulses, frequency step size, and pulse width. The previous sections addressed some of these relationships. The principal trades are summarized here for convenience:

1. Pulse repetition interval, T :
 - a. The minimum PRI is selected to support the maximum unambiguous range.
 - b. The product of the PRI and the number of pulses, NT , defines the dwell time associated with the waveform. The dwell time should be sized to prevent a target from moving through more than $1/2$ a range resolution cell during a coherent processing interval.
2. Product of the number of pulses and the frequency step size, $N\Delta f$:
 - a. The product defines the waveform's composite bandwidth.
 - b. The radar system's range resolution, $c/2N\Delta f$, dictates the value of $N\Delta f$.
3. Frequency step size, Δf :
 - a. Defines the size of the unambiguous region, $c/2\Delta f$.
 - b. A small value of Δf increases the number of pulses required to achieve the desired range resolution and correspondingly the waveform's dwell time.
 - c. When imaging a target of length L , $c/2\Delta f > L$ is often enforced to resolve ambiguous returns.
 - d. In many cases, $c/2\Delta f$ defines the range gate spacing.
4. Pulse width, τ (CW pulse only):
 - a. With $\tau = 2L/c$, the maximum straddle loss is 6 dB across a target of length L .
 - b. With $\tau = 1/2\Delta f$, returns are localized to a region defined by $c/2\Delta f$; however, straddle loss tends to increase.

2.5.6 Impacts of Doppler

The SF waveform has been examined assuming a stationary target and radar. This is a valid assumption in some instances, for example, an instrumentation radar used to measure the HRR profile of a stationary object. In other instances (e.g., a tactical application), the environment is not as constrained, and the effects of relative motion must be considered.

To evaluate the impact of Doppler shift on an SF waveform, first consider the samples collected from a stationary target as defined in equation (2.112). With relative motion, the measured range to the target changes pulse to pulse. The range varies as

$$R(n) = R_0 - nvT \quad n = 0 \dots N - 1 \quad (2.129)$$

where v is the radial component of velocity.

Inserting (2.129) into equation (2.112)

$$\hat{x}(n) = \exp\left(j2\pi(f_0 + n\Delta f)\frac{2(R_0 - nvT)}{c}\right) \quad n = 0, \dots, (N-1) \quad (2.130)$$

and expanding terms

$$\hat{x}(n) = \exp\left(j2\pi\frac{2R_0}{c}f_0\right) \exp\left(j2\pi\frac{2R_0}{c}n\Delta f\right) \exp\left(-j2\pi\frac{2nvT}{c}f_0\right) \exp\left(-j2\pi\frac{2n^2vT}{c}\Delta f\right) \quad (2.131)$$

The first exponential in (2.131) is a constant phase term. The second exponential contains a linear phase term associated with the range to the target. Both of these terms are present in the zero Doppler case (in equation (2.112)), and the remaining two exponentials are a direct result of the relative motion between the radar and the target. A Doppler shift is also imparted to the carrier frequency, but the associated phase contribution is small compared with the values defined in equation (2.131) and is therefore ignored.

The third term in equation (2.131) introduces range-Doppler coupling with the peak of the response shifted in range by

$$r_{shift} = \frac{vTf_0}{\Delta f} \quad (2.132)$$

Normalizing the range shift in equation (2.132) by the waveform's range resolution yields

$$\bar{r}_{shift} = \frac{2NvTf_0}{c} \quad (2.133)$$

which expresses the shift in terms of range resolution cells. The fourth exponential is quadratic in n producing a spreading in the compressed response. The maximum spread occurs when $n = N$ or

$$r_{spread} = NvT \quad (2.134)$$

Normalizing the spread in equation (2.134) by the waveform's range resolution yields

$$\bar{r}_{spread} = \frac{2vN^2T\Delta f}{c} \quad (2.135)$$

The spreading term produces a loss in coherent gain and thus a loss in SNR. The range-Doppler response of a stepped frequency waveform is very similar to an LFM waveform with exception that the response aliases when the Doppler shift causes a range displacement greater than $c/2\Delta f$.

For a single target, the effects of Doppler may be mitigated if an estimate of the target's radial velocity is ascertained via tracking or some other means. A range-independent correction factor may be applied to the samples prior to applying a DFT:

$$x_{correct}(n) = \exp\left(j2\pi\frac{2n\hat{v}T}{c}f_0\right) \exp\left(j2\pi\frac{2n^2\hat{v}T}{c}\Delta f\right) \quad (2.136)$$

where \hat{v} is an estimate of the radial velocity. The correction factor is multiplied by the measured samples in equation (2.131) to remove the undesired terms.

2.5.7 Summary

An SF waveform is designed to achieve a narrow instantaneous bandwidth on both transmit and receive to accommodate hardware limitations. Fortunately, a wide bandwidth signal may be synthesized by the stepping pulses over frequency and coherently processing the returns. The achieved range resolution is inversely proportional to the total frequency extent over which the waveform is stepped. The waveform requires a long collection interval, which often limits its application; however, the waveform is used in some wide-band instrumentation radars and in some tactical systems. The frequency stepping introduces ambiguities in range that must be addressed through judicious waveform parameter selection.

2.6 | QUADRIPHASE SIGNALS

2.6.1 Introduction

Biphase-coded waveforms are formed by concatenating chips (or subpulses) and modulating the phase chip to chip. The phase of each chip is either 0 or 180 degrees. The spectrum of the resultant waveform retains many of the characteristics of the constituent subpulse including high spectral sidelobes. The sidebands of the spectrum are a source of EMI and have the potential to interfere with other RF systems operating nearby. Quadrphase waveforms were developed by Taylor and Blinchikoff [27] for use in the AN/TPS-63 radar to reduce the spectral sidelobes of biphase-coded waveforms while retaining their time-domain sidelobe performance [28-30]. Quadrphase encoding is similar to the minimum shift keying (MSK) technique applied in some communication systems and may also be viewed as a modified version of frequency shift keying (FSK) [31].

The relatively high spectral sidelobes, characteristic of phase-coded signals, are an artifact of the rectangular subpulses commonly used to synthesize the waveform. A nonrectangular-shaped subpulse (e.g., a Gaussian-shaped subpulse) may be used to suppress the sidelobes, but the waveform's envelope exhibits a time-varying amplitude response that is undesirable from an energy perspective. Taylor and Blinchikoff [27] developed a biphase-to-quadrphase (BTQ) transformation that produces a waveform with the following properties:

1. Spectral sidelobes lower than those achieved with a phase-coded waveform employing rectangular subpulses
2. Time sidelobes near those achieved with a specified biphase-coded waveform
3. A nearly constant amplitude response

2.6.2 BTQ Transformation

The BTQ transformation may be applied to any biphase code. The first step is to select a biphase code that exhibits the desired range sidelobe response. Next, the biphase code is transformed into a four-phase (quadrphase) code:

$$q_n = j^{s(n-1)} c_n \quad n = 1, 2, \dots, N \quad (2.137)$$

TABLE 2-3 ■ Conversion of a 13-Element Barker Code into a Quadriphase Code

Element Index, n												
1	2	3	4	5	6	7	8	9	10	11	12	13
13-Element Barker Code												
1	1	1	1	1	-1	-1	1	1	-1	1	-1	1
13-Element Quadriphase Code												
1	j	-1	$-j$	1	$-j$	1	$-j$	1	$-j$	-1	j	1

where c_n is the original biphase code and $c_n \in \{1, -1\}$, q_n is the resultant quadriphase code, n is the chip index, s is either 1 or -1 , and N is the length of the biphase code. Since c_n is biphase, the elements of q_n are real for odd values of n and imaginary for even values of n . Table 2-3 contains the transformation for a 13-element biphase Barker code.

The transformation in equation (2.137) is not sufficient by itself to create the desired spectral response. The next step is to define to a subpulse, $p(t)$, whose shape is a half-cosine

$$p(t) = \cos(\pi t / 2\tau_c) \quad -\tau_c \leq t \leq \tau_c \quad (2.138)$$

where the quadriphase subpulse width is $2\tau_c$, and τ_c is the chip width associated with the biphase-coded waveform.

A complex signal is formed by summing weighted and shifted copies of the subpulse

$$z(t) = \sum_{n=1}^N j^{s(n-1)} c_n p(t - n\tau_c) \quad (2.139)$$

where adjacent subpulses overlap in time by τ_c seconds. The real and imaginary parts of the complex signal formed using a 13-element Barker are plotted in Figure 2-32. An important feature of the quadriphase signal is that the envelope

$$a(t) = |z(t)| \quad (2.140)$$

is constant over most of the response, as shown in Figure 2-33. The only exceptions are the segments at the beginning and end of the waveform. The phase of the complex signal

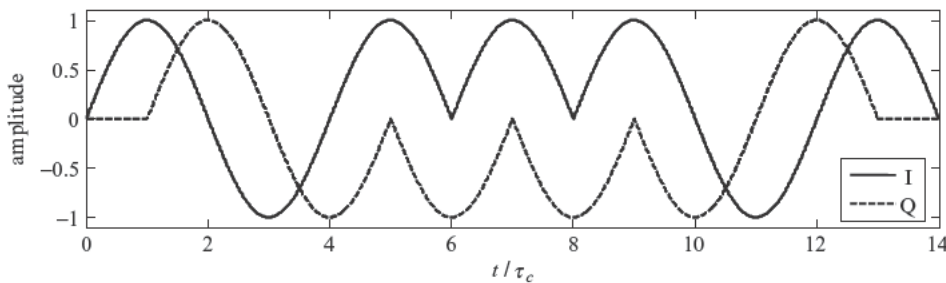
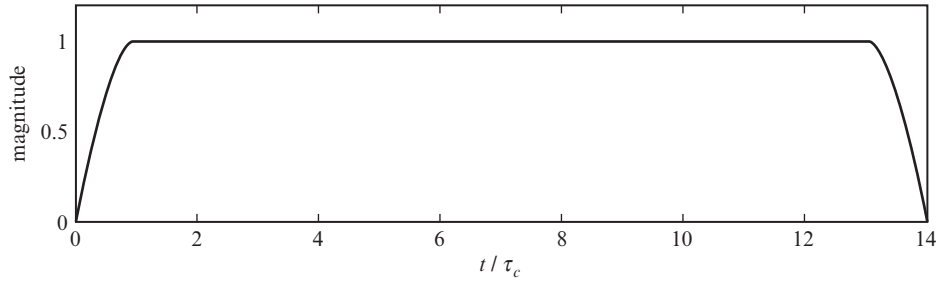
**FIGURE 2-32** ■ The quadriphase transformation converts a biphase code into a complex signal. The chips consist of half-cosine subpulses, and the phase is continuously varying.

FIGURE 2-33 ■ The envelope of the complex signal is constant except at the ends of the pulse.



in equation (2.139) is

$$\phi(t) = \tan^{-1} \left(\frac{\text{Im}\{z(t)\}}{\text{Re}\{z(t)\}} \right) \quad (2.141)$$

and is plotted in Figure 2-34. Over a chip, the phase of the signal varies linearly, whereas the phase of a traditional phase-coded waveform is constant.

2.6.3 Quadriphase Waveform

In a radar system, the amplitude and phase of the complex signal in equation (2.139) are used to modulate a sinusoid

$$y(t) = a(t)\cos(\phi(t)) \quad (2.142)$$

The signal in equation (2.142) is centered at baseband. The signal is then mixed to an RF and transmitted.

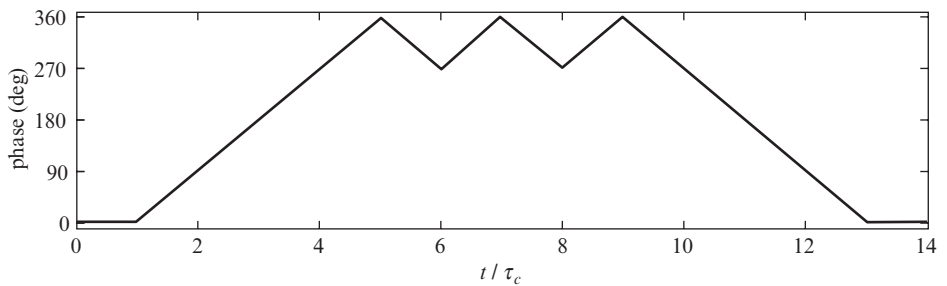
2.6.4 Waveform Properties

2.6.4.1 Spectrum

With a phase-coded waveform, the chip's shape determines the envelope of the waveform's spectrum [1]. A similar relationship holds for quadriphase waveforms. Taylor and Blinchikoff [27] show that the spectrum of a half-cosine pulse is

$$|P(\Omega)|^2 = \left(\frac{4\tau_c}{\pi} \right)^2 \frac{\cos^2(\Omega\tau_c)}{\left(1 - 16 \left(\frac{\Omega\tau_c}{2\pi} \right)^2 \right)^2} \quad (2.143)$$

FIGURE 2-34 ■ The phase of the quadriphase code varies linearly across a chip.



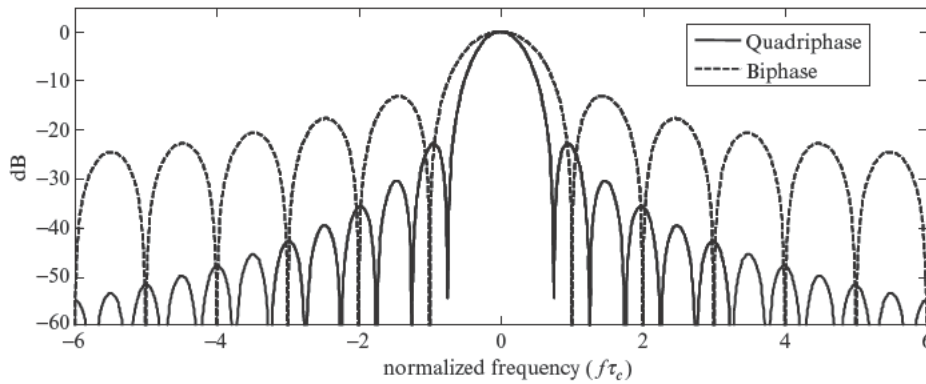


FIGURE 2-35 ■ The subpulse shape determines the basic envelope of the waveform's spectrum.

Figure 2-35 contains a comparison of the spectrum of a half-cosine pulse of length $2\tau_c$ and the spectrum of a rectangular pulse of length τ_c . The half-cosine pulse exhibits lower sidelobes that roll off at a faster rate than the rectangular pulse.

The actual spectrum of a quadriphase waveform is a function of both the subpulse and the quadriphase code. The waveform may be expressed as a convolution of the quadriphase code, with elements spaced by τ_c , and the half-cosine pulse of length $2\tau_c$

$$z(t) = p(t) * \sum_{n=1}^N q_n \delta(t - n\tau_c) \quad (2.144)$$

where $*$ denotes linear convolution. The waveform's spectrum is the product of the half-cosine pulse's spectrum in equation (2.143) and the spectrum of the quadriphase code. Figure 2-36 contains a comparison the spectrum associated with a 13-element Barker code and its corresponding quadriphase code. The quadriphase waveform has a narrower bandwidth, corresponding to a loss in resolution, but has substantially lower spectral sidelobes. The peak spectral sidelobe of the quadriphase waveform is approximately -23 dB, compared with approximately -13 dB for a traditional phase code with rectangular subpulses. The spectral sidelobes of the quadriphase code roll off at 12 dB per octave, whereas the sidelobes of the sinc response roll off at 6 dB per octave. The BTQ transformation has achieved the desired goal of reducing the energy in the sidelobes and thus the effects of EMI. The impact on the match filtered response is examined in the next section.

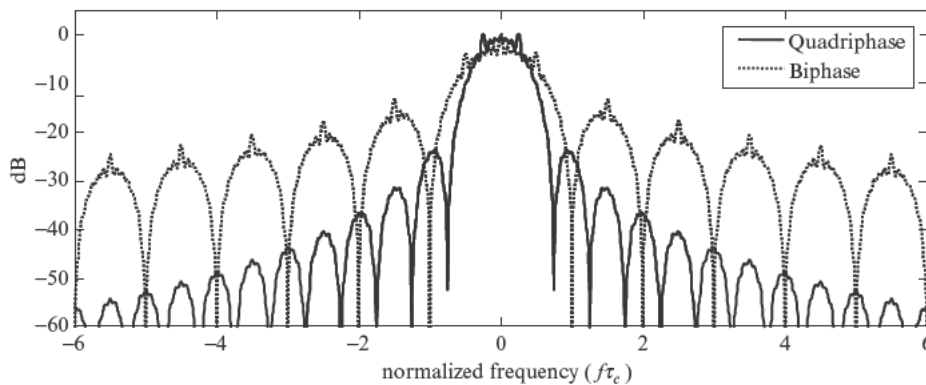
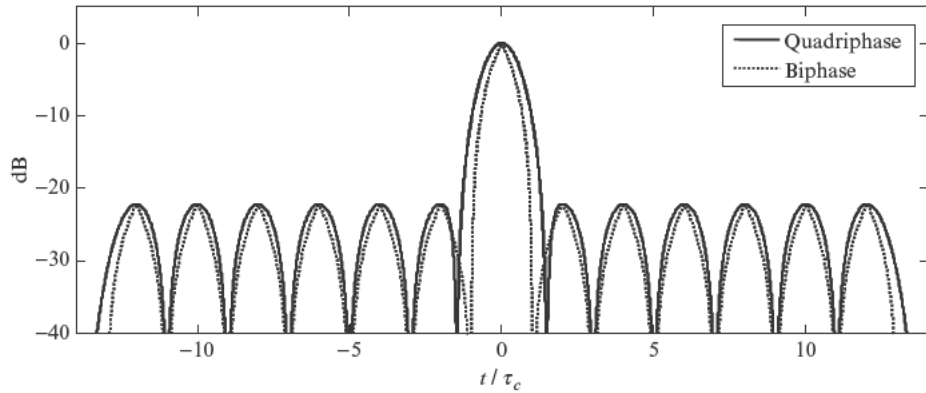


FIGURE 2-36 ■ The waveform's spectrum is the product of the subpulse's and code's spectra.

FIGURE 2-37 ■ The quadriphase code exhibits peak sidelobes equal to the 13-element Barker code. The main lobe of the quadriphase code is broadened, resulting in a loss of resolution.



2.6.4.2 Match Filtered Response

As previously noted, the main lobe of the quadriphase spectrum in Figure 2-36 is narrower than that of the biphase code, corresponding to a slight reduction in bandwidth and manifesting as a loss in resolution. The match filtered responses for the quadriphase and biphase-coded waveforms are plotted in Figure 2-37. The quadriphase waveform exhibits a wider main lobe and thus a loss in range resolution. Taylor and Blinchikoff [27] predict the broadened response by examining the half-cosine's autocorrelation response

$$m(t) = \tau_c \left(1 - \frac{|t|}{2\tau_c} \right) \cos \frac{\pi t}{2\tau_c} + \frac{\tau_c}{\pi} \sin \frac{\pi |t|}{2\tau_c} \quad -2\tau_c \leq t \leq 2\tau_c \quad (2.145)$$

The main lobe width measured at the 6 dB point is approximately $1.54 \tau_c$ compared with τ_c for a rectangular subpulse. The corresponding 3 dB points are $1.1 \tau_c$ and $0.58 \tau_c$ for the quadriphase and rectangular subpulses, respectively.

For an ideal biphase-coded waveform with rectangular chips, the sidelobes are defined entirely by the code sequence. With Taylor and Blinchikoff's [27] quadriphase waveform, adjacent subpulses overlap resulting in cross-talk and a slight degradation in sidelobe performance. The ratio of the quadriphase peak sidelobe to the biphase peak sidelobe is

$$QBR = \sqrt{1 + \left(\frac{2}{\pi} \right)^2 \left(1 - \frac{1}{a} \right)^2} \quad (2.146)$$

where a is the peak sidelobe of the biphase code. For Barker codes, the peak sidelobe is 1 and thus $QBR = 1$ (i.e., the peak sidelobes of the biphase and quadriphase codes are identical). For non-Barker codes, the ratio may approach a maximum value of 1.185 (or 1.48 dB).

2.6.4.3 Ambiguity Surface

While the sidelobe structure of a quadriphase waveform is quite similar to its parent biphase code, Levanon and Freedman [30] show that its response in the presence of a Doppler shift is quite different. A quadriphase code derived using a 13-element Barker exhibits a range-Doppler coupled ridge similar to an LFM waveform, whereas the biphase code exhibits a distorted thumbtack response. In general, the ambiguity surface of a quadriphase code may be quite different from that of the parent biphase code.

2.6.5 Summary

Quadrphase codes achieve lower spectral sidelobes than traditional phase codes employing rectangular subpulses. The lower sidelobes reduce the potential for EMI. Quadrphase codes are generated by applying a transformation to the biphasic code and using subpulses shaped as half-cosines. The subpulses are overlapped to create a nearly constant envelope. The peak sidelobe performance is at or near that achieved by the biphasic code. At most, the peak sidelobe increases by 1.48 dB. At the 3 dB point, the main lobe increases by a factor 1.9 corresponding to a loss in range resolution. The ambiguity surface of a quadrphase code may be quite different from that associated with the parent biphasic code.

2.7 | MISMATCHED FILTERS

2.7.1 Introduction

Phase codes exhibit a sidelobe response defined by the code sequence. MMFs may be applied to reduce or shape the sidelobes at the expense of a slight loss in SNR. This section examines the design of mismatched filters applied to a phase-coded waveform to tailor the sidelobes.

2.7.2 Performance Metrics

When evaluating mismatched filter performance, three metrics are often used: PSR; integrated sidelobe ratio (ISR); and loss in processing gain (LPG). PSR is defined as the square of the largest sidelobe divided by the square of the main lobe peak. ISR is the energy in the sidelobes (i.e. the sum of the squares of all the sidelobes) normalized by the square of the main lobe peak. LPG is the ratio of the square of the main lobe peak when employing mismatched filtering to the square of the main lobe peak when employing matched filtering and is equivalent to the loss in SNR.

2.7.3 Mismatched Filter Approaches

MMFs may be designed to minimize the ISR, to reduce the PSR, or to tailor specific regions of the sidelobe response. The process of creating a filter is examined in the following sections.

2.7.3.1 Minimum ISR Filter

The lowest ISR known for a binary sequence is -11.49 dB, which is achieved by a 13-element Barker code; for longer binary sequences, the lowest achieved ISR is approximately -8 dB [43, 44]. A mismatched filter may be applied that achieves a lower ISR and that in many cases also yields a lower PSR. The filter is derived as the least squares solution to an overdetermined set of equations used to model a desired sidelobe response.

Consider a phase code c_k , where $k = 0, \dots, K - 1$, and a FIR filter with coefficients z_m where $m = 0, \dots, M - 1$ and $M \geq K$. The filtered response is

$$y_n = \sum_{k=0}^{K-1} c_k z_{n-k} \quad n = 0, \dots, (K + M - 2) \quad (2.147)$$

The desired filtered response is d_n where

$$d_{n=p} = 1 \quad (2.148)$$

and

$$d_n = 0 \quad n \neq p \quad (2.149)$$

where $p = (M + K)/2$ for $M + K$ an even number, and $p = (M + K - 1)/2$ for $M + K$ an odd number. The error between the desired response and the response achieved using the filter is

$$e_n = d_n - y_n \quad (2.150)$$

and the sum of the squared errors is

$$E = \sum_{n=0}^{K+M-2} (d_n - y_n)^2 \quad (2.151)$$

The filter operation in equation (2.147) may be written using matrix notation

$$\mathbf{y} = \mathbf{C}\mathbf{z} \quad (2.152)$$

where \mathbf{y} is a column vector containing the filtered response with dimension $(K + M - 1) \times 1$, \mathbf{C} is a matrix with dimension $(K + M - 1) \times M$ whose rows contain delayed copies of the phase code, and \mathbf{z} is a column vector containing the filter coefficients. An example of the \mathbf{C} matrix is

$$\mathbf{C} = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ c_2 & c_1 & \cdots & 0 \\ 0 & \cdots & c_K & c_{K-1} \\ 0 & \cdots & 0 & c_K \end{bmatrix} \quad (2.153)$$

The sum of the squared error may be expressed using matrix notation

$$E = (\mathbf{d} - \mathbf{y})(\mathbf{d} - \mathbf{y})^H \quad (2.154)$$

where \mathbf{d} is a vector with dimension $(K + M - 1) \times 1$ containing the desired response, and H denotes Hermitian transpose. A least squares (LS) solution [44] to the overdetermined set of equations may be obtained which minimizes the squared error and thus the ISR. The LS filter coefficients are

$$\mathbf{z} = (\mathbf{C}^H \mathbf{C})^{-1} \mathbf{C}^H \mathbf{d} \quad (2.155)$$

2.7.3.2 Minimum ISR Example

A few examples are used to examine minimum ISR (MISR) filters and their properties. A 27-element MISR filter is generated using a 13-element Barker code. The compressed responses for a 13-element Barker having applied a matched filter and a 27-element MISR filter are plotted in Figure 2-38. As previously noted, this particular Barker code exhibits a -11.5 dB ISR. Using the MISR filter, the ISR is reduced by 8.3 dB, and the PSR is reduced by approximately 6 dB. The LPG is 0.18 dB, which is small compared with the loss associated with a typical weighting function.

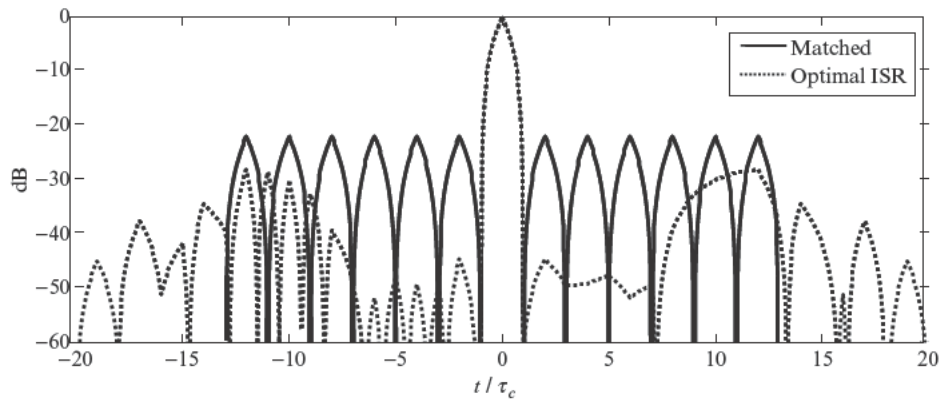


FIGURE 2-38 ■ The optimal ISR filter exhibits a lower PSR and ISR than the match filtered 13-element Barker code.

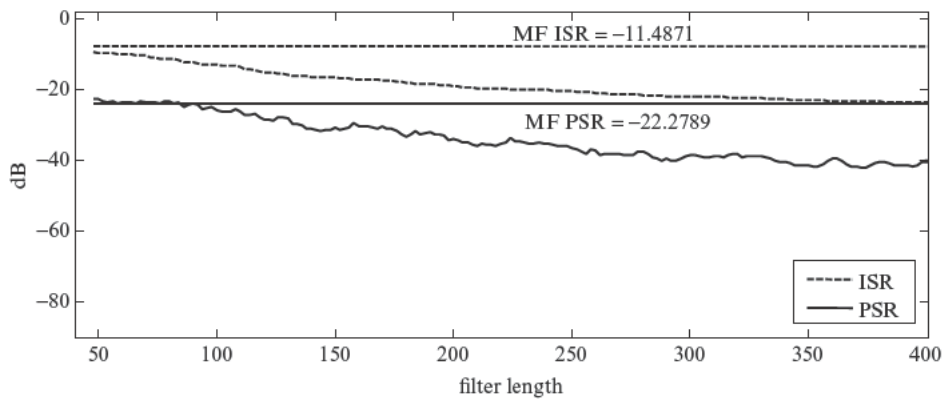


FIGURE 2-39 ■ The PSR and ISR decrease with increasing filter length. In this example, an MISR is applied to a 48-element MPS code.

The filter length may be increased to achieve lower sidelobes, but the link between sidelobe suppression and filter size is code dependent. Codes with a low match filtered PSR generally exhibit a more rapid reduction in sidelobes with increasing filter length, but even codes with similar match filtered characteristics exhibit varying degrees of performance. Searches [33, 46, 47] may be performed to identify codes that perform well with MMFs. In many instances, these codes exhibit better MMF sidelobe performance than codes selected solely on the basis of their match filtered properties.

Figure 2-39 contains the ISR and PSR for a 48-element minimum peak sidelobe (MPS) code [48] as a function of filter length. The 48-element MPS code is provided in Table 2.4. Applying a 400-element MISR filter, the LPG is approximately 1.5 dB. A biphasic code optimized to exhibit low sidelobes when combined with MMF filtering is provided in Table 2-4. Figure 2-40 contains a plot of the code's performance as a function of filter length. Applying a 400-element filter, the LPG is approximately 0.86 dB.

TABLE 2-4 ■ Phase Codes Used in Mismatched Filter Examples

Code Type	Elements (Octal Representation)
48-element MPS code	0526554171447763
48-element optimized code for MMF	2522332235416377
64-element MPS code	0061243076734133115722

FIGURE 2-40 ■ A MISR filter is applied to a 48-element code optimized to perform well with mismatched filtering. The PSR and ISR are significantly reduced compared with the 48-element MPS code employing an MISR filter.

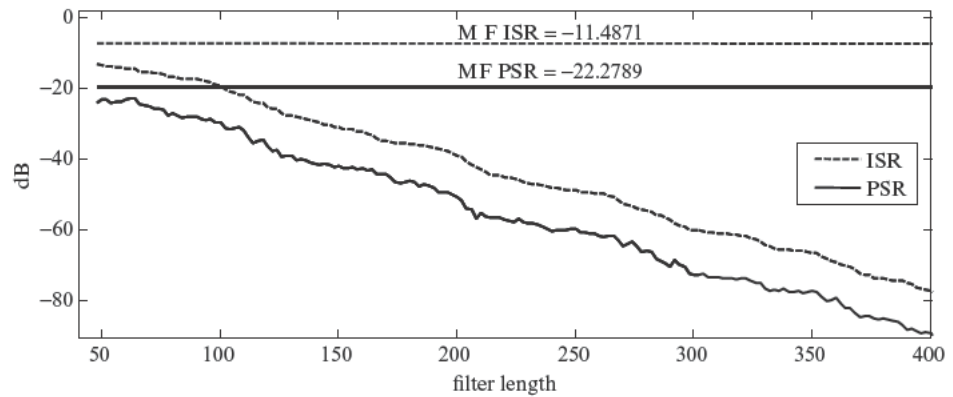


TABLE 2-5 ■ Performance of MMFs Applied to a 48-Element MPS Code and an Optimized Code

	48-Element MPS Code			48-Element Optimized Code for MMF		
	LPG (dB)	ISR (dB)	PSR (dB)	LPG (dB)	ISR (dB)	PSR (dB)
MF	0.00	-7.9	-24.1	0.00	-7.35	-19.6
MMF-100	0.66	-13.2	-26.2	0.79	-19.3	-29.5
MMF-200	1.10	-19.2	-34.1	0.86	-39.2	-50.5
MMF-300	1.32	-22.1	-38.6	0.86	-60.0	-72.5
MMF-400	1.47	-23.9	-40.6	0.86	-77.0	-89.0

The performance achieved using different filter lengths is summarized in Table 2-5. The first set of metrics (LPG, ISR, and PSR) is associated with the MPS code and the second set with the optimized code. Although the MPS code has better matched filter performance, the optimized code has better mismatched filter performance.

2.7.3.3 Doppler Tolerance

The performance achieved with mismatched filtering degrades quickly in the presence of a Doppler shift. Figure 2-41 shows the effect of Doppler on the optimized 48-element sequence and a 250-element MISR filter. The horizontal axis defines the number of cycles, $f_d \tau$, across the uncompressed pulse length, τ . The PSR and ISR degrade quickly even with only a 1/100 cycle of Doppler over the uncompressed pulse. In general, MMF performance is very sensitive to Doppler.

2.7.3.4 Tailoring the Sidelobe Response

The filter in equation (2.155) may be generalized to tailor the sidelobe response. A weighted least squares solution provides a mechanism for assigning a relative weight to each sidelobe. Larger weights indicate a higher degree of importance in meeting the desired sidelobe response, whereas smaller weights imply less importance. The weighted least squares filter is

$$\mathbf{z} = (\mathbf{C}^H \mathbf{W} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{d} \quad (2.156)$$

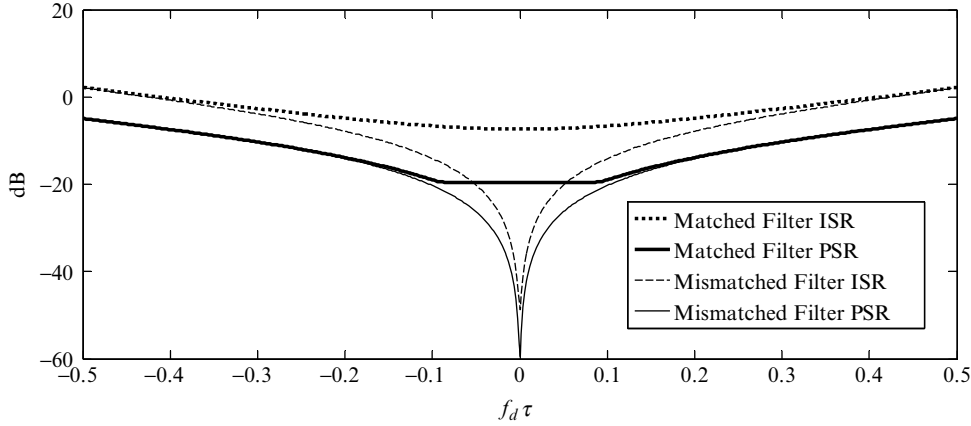


FIGURE 2-41 ■ A small fractional Doppler shift has a dramatic impact on MMF performance. The plot shows the degraded performance of the optimized 48-element code and a 250-element MISR filter as a function of the number of cycles of Doppler shift across a pulse.

where W is a diagonal weighting matrix whose elements are the weights assigned to each sidelobe

$$W = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & w_{K+M-2} & 0 \\ 0 & 0 & 0 & 0 & w_{K+M-1} \end{bmatrix} \quad (2.157)$$

and $w > 0$.

For a given code and filter length, the weighted LS (WLS) solution does not achieve an ISR lower than the MLS solution, but the sidelobe energy is distributed in a desired manner. Consider an application where near-in sidelobes (i.e., those closest to the main lobe) are more detrimental than distant sidelobes. To create a response with low near-in sidelobes, W is defined with the larger weights assigned to the near-in sidelobes. The filter will suppress the near-in sidelobes at the expense of higher sidelobes elsewhere. A filter used to create this type of response is termed a null-region filter.

Figure 2-42 contains the compressed responses for a 130-element null-region filter and an MISR filter both derived for a 64-element MPS code [49]. The code is provided in Table 2-4. For the null-region filter, the 40 sidelobes on either side of the main lobe are assigned a weight of 10, and the remaining sidelobes are assigned a weight of 1. The filter suppresses the near-in sidelobes to a peak value of approximately -70 dB. The remaining sidelobes approach the levels achieved with the MISR filter. For the null-region filter, ISR and PSR increase by 2.2 dB and 4.6 dB, respectively, and the LPG increases by 0.63 dB. By adjusting the weights assigned to the near-in sidelobes, it is possible to find a balance between the suppression achieved in the null region and other performance metrics.

2.7.3.5 Reduced Peak Sidelobe Filters

Mismatched filters may also be designed to reduce the PSR to a desired level. Such a filter may be obtained using linear programming [32] or through an iterative process [34]

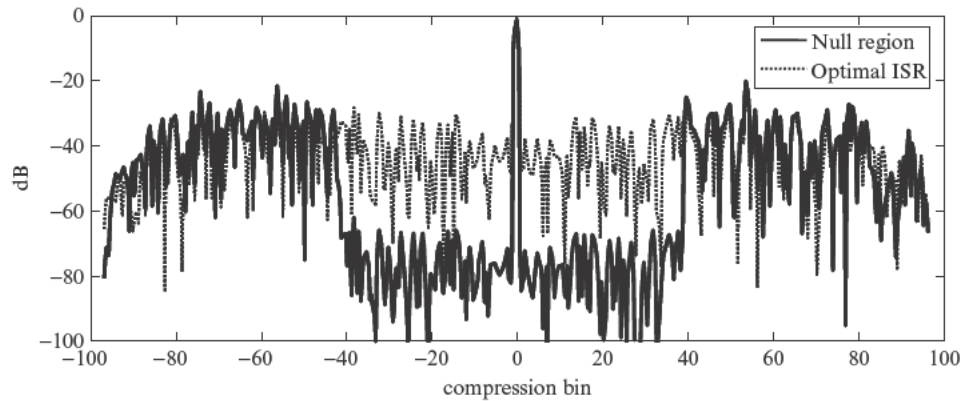


FIGURE 2-42 ■ A WLS filter may be used to create a null-region filter that suppresses the near-in sidelobes. The null-region filter is applied to a 64-element MPS code and is designed to suppress the first 40 bins on either sidelobe of the main lobe.

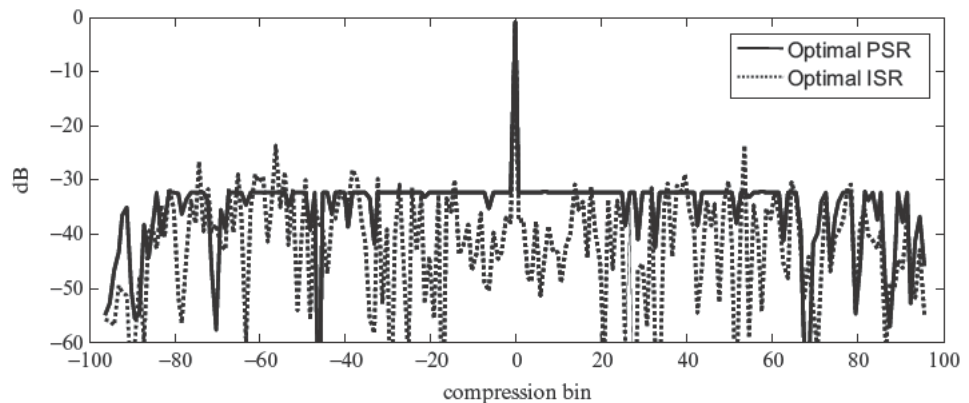
employing the WLS equations. The steps involved in the iterative WLS approach are as follows:

1. The MISR filter is obtained using equation (2.155) for a given phase code and filter length.
2. The weighted LS solution is obtained using equation (2.156). The weights are assigned the magnitude of the sidelobes obtained for the filtered response in step 1. A small offset is added to each weight to prevent any zero or near-zero values.
3. Step 2 is repeated until the filter achieves a peak sidelobe level within some tolerance of the desired value.

A filter obtained via this approach will tend to suppress the higher sidelobes, and if repeated a number of times the sidelobes will approach a flat response.

The compression of a 64-bit MPS code (defined in Table 2-4) using a 130-element reduced PSR filter is shown in Figure 2-43. The filter reduces the PSR by 8.3 dB; however, the ISR increases by 2.3 dB, and the LPG increases by 0.26 dB. A number of the sidelobes are higher than those associated with the optimal ISR filter, yet the PSR is reduced. As the peak sidelobe is reduced, the energy in the sidelobe region is distributed more evenly among the remaining sidelobes. The observed ISR is equal to or greater than that achieved with the optimal ISR filter.

FIGURE 2-43 ■ The peak sidelobes may be reduced through an iterative process. A reduced PSR filter of length 130 is applied to a 64-element MPS code.



2.7.4 Summary

While matched filters provide the maximum SNR for a given code length, MMFs are capable of providing substantially lower sidelobes with only a slight increase in SNR loss. MMFs may be designed to minimize ISR, to reduce PSR, and to tailor the sidelobe response. Extremely low sidelobes are possible, but MMFs are not Doppler tolerant. A small fractional Doppler shift quickly degrades sidelobe performance.

2.8 | FURTHER READING

The references serve as a good resource for further reading. Stretch processing is covered in radar and waveform texts [3, 4] and in several SAR texts [5, 6, 40]. Stepped chirp waveforms are relatively new and are not addressed in most radar texts. The reader is encouraged to start with the papers by Lord and others [7, 8, 10–13]. The initial work in NLFM design performed by Key and others [19–21] is summarized in two excellent texts [26, 50]. Papers by Johnston [24, 25] are also recommended for those interested in the Doppler intolerance associated with NLFM waveforms.

Stepped frequency waveforms are presented in a coherent fashion in the textbook by Wehner [18] and in the chapter by Lane [17]. The original paper by Taylor and Blinchikoff [27] provides background on the motivation behind the development of the quadriphase waveform as well as derivations for some of the equations in this text. Levanon and Mozeson [51] describe other band-limiting techniques, in addition to quadriphase, and describe the ambiguity surface for a quadriphase waveform. The various MMF designs are covered at length in [32–35]. Early approaches to MMF design are covered in [52, 53], and more recently an adaptive approach was described in [54].

2.9 | REFERENCES

- [1] Keel, B.M., “Fundamentals of Pulse Compression Waveforms,” Chapter 20 in *Principles of Modern Radar, Basic Principles*, Ed. M. A. Richards, J. A. Scheer, and W. A. Holm, Scitech Publishing, Raleigh, NC, 2010.
- [2] Caputi, Jr., W.J., “Stretch: A Time-Transformation Technique,” *IEEE Trans. AES*, vol. AES-7, no. 2, pp. 269–278, March 1971.
- [3] Skolnik, M., *Radar Handbook*, 3d ed., McGraw-Hill, New York, 2008.
- [4] Stimson, G.W., *Introduction to Airborne Radar*, 2d ed., Scitech Publishing, Raleigh, NC, 1998.
- [5] Jakowatz, Jr., C.V., Wahl, D.E., Eichel, P.H., Ghiglia, D.C., and Thompson, P.A., *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Kluwer Academic Publishers, Boston, 1996.
- [6] Curlander, J.C. and McDonough, R.N., *Synthetic Aperture Radar, Systems and Signal Processing*, New York, John Wiley & Sons, 1991.
- [7] Nel, W., Tait, J., Lord, R., and Wilkinson, A., “The Use of a Frequency Domain Stepped Frequency Technique to Obtain High Range Resolution on the CSIR X-Band SAR System,” in *Proceedings of the IEEE 2002 Africon Conference*, vol. 1, pp. 327–332, 2002.

- [8] Wilkinson, A.J., Lord, R.T., and Inggs, M.R., "Stepped-Frequency Processing by Reconstruction of Target Reflectivity Spectrum," in *Proceedings of the 1998 South African Symposium on Communications and Signal Processing*, pp. 101–104, 1998.
- [9] Berens, P., "SAR with Ultra-High Range Resolution Using Synthetic Bandwidth," in *Proceedings of the IEEE 1999 International Geoscience and Remote Sensing Symposium*, vol. 3, pp. 1752–1754, 1999.
- [10] Lord, R.T. and Inggs, M.R., "High Resolution SAR Processing Using Stepped-Frequencies," in *Proceedings of the 1997 IEEE International Geoscience and Remote Sensing Conference*, pp. 490–492, August 1997.
- [11] Lord, R.T. and Inggs, M.R., "High Range Resolution Radar using Narrowband Linear Chirps Offset in Frequency," in *Proceedings of the 1997 South African Symposium on Communications and Signal Processing*, pp. 9–12, September 1997.
- [12] Hongxing, D., "Stepped Frequency Chirp Signal SAR Imaging," in *Proceedings of the First Asian and Pacific Conference on Synthetic Aperture Radar*, pp. 14–18, 2007.
- [13] Haojuan, Y., Meiguo, G., and Guoman, L., "Coherent Spectrum Synthesis of Frequency-Stepped Chirp Signal," in *Proceedings of the 2009 IET International Radar Conference*, pp. 1–4, April 2009.
- [14] McGroary, F.X., Lindell, K., and Greenspan, M., "Multi-Pulse Pulse Compression Radar System," US Patent Number 5,128,681, July 7, 1992.
- [15] McGroary, F. and Lindell, K., "A Stepped Chirp Technique for Range Resolution Enhancement", in *Proceedings of the National Telesystems Conference*, pp. 122–126, 1991.
- [16] Ruttenberg, K. and Chanzit, L., "High Range Resolution by Means of Pulse to Pulse Frequency Shifting," in *Proceedings of the EASCON, Radar Systems*, vol. 2, pp. 153–157, 1968.
- [17] Lane, T. L., "Stepped-Frequency and ISAR Imaging Systems," Chapter 11, in *Coherent Radar Performance Estimation*, Ed. J. A. Scheer and J. L. Kurtz, Artech House, Norwood, MA, 1993.
- [18] Wehner, D. R., "Synthetic High-Range-Resolution Radar", Chapter 5, in *High-Resolution Radar*, 2d ed., Artech House, Boston, MA, 1995.
- [19] Key, E.L., Fowle, E.N., and Haggarty, R.D., "A Method of Designing Signals of Large Time-Bandwidth Product," in *Proceedings of the IRE International Convention Record*, vol. 9, pt. 4, pp. 146–154, 1961.
- [20] Fowle, E.N., "The Design of FM Pulse Compression Signals," *IEEE Trans. on Information Theory*, pp. 61–67, January 1964.
- [21] Newton, C.O., "Nonlinear Chirp Radar Signal Waveforms for Surface Acoustic Wave Pulse Compression Filters," *Wave Electronics*, pp. 387–401, 1974.
- [22] Brandon, M.A., "The Design of a Non-linear Pulse Compression System to Give a Low Loss High Resolution Radar Performance," *Marconi Review*, vol. 36, no. 188, pp. 1–45, 1973.
- [23] Brandon, M.A., "The Spectra of Linear and Non-linear F.M. Used in Pulse Compression, and the Effects on the Resultant Compressed Pulse," *Marconi Review*, vol. 36, no. 188, pp. 69–92, 1973.
- [24] Johnston, J.A., "Analysis of the Pulse Compression of Doppler Shifted Nonlinear Frequency Modulated Signals," *Electronic Letters*, vol. 20, pp. 1054–1055, December 1984.
- [25] Johnston, J.A. and Fairhead, A.C., "Waveform Design and Doppler Sensitivity Analysis for Nonlinear FM Chirp Pulses," *IEE Proceedings-F*, vol. 133, no. 2, pp. 163–175, April 1986.
- [26] Cook, C.E. and Bernfeld, M., *Radar Signals: An Introduction to Theory and Application*, Artech House, Boston, 1993.

- [27] Taylor, Jr., J.W. and Blinchikoff, H.J., “Quadrphase Code—A Radar Pulse Compression Signal with Unique Characteristics,” *IEEE Trans. on AES*, vol. 24, no.2, pp. 156–170, March 1988.
- [28] Faust, H., Connolly, B., Firestone, T.M., Chen, R.C., Cantrell, B.H., and Mokole, E.L., “A Spectrally Clean Transmitting System for Solid-State Phased-Array Radars,” in *Proceedings of the IEEE Radar Conference*, pp. 140–144, April 26–29, 2004.
- [29] Blinkchikoff, H.J., “Range Sidelobe Reduction for the Quadrphase Codes,” *IEEE Trans. AES*, vol. 32, no. 2, pp. 668–675, April 1996.
- [30] Levanon, N. and Freedman, A., “Ambiguity Function of Quadrphase Coded Radar Pulse,” *IEEE Trans. AES*, vol. 25, no. 6, pp. 848–853, November 1989.
- [31] Vale, C.R., “SAW Quadrphase Code Generator,” *IEEE Trans. on Sonics and Ultrasonics*, vol. SU-28, no. 3, pp. 132–136, May 1981.
- [32] Stoica, P., Li, J., and Xue, M., “Transmit Codes and Receive Filters for Pulse Compression Radar Systems,” in *Proceedings of the IEEE 2008 International Conference on Acoustics, Speech and Signal Processing*, pp. 3649–3652, March 31–April 4, 2008.
- [33] Nunn, C. and Kretschmer, F.F., “Performance of Pulse Compression Code and Filter Pairs Optimized for Loss and Integrated Sidelobe Level,” in *Proceedings of the 2007 IEEE Radar Conference*, pp. 110–115, April 17–20, 2007.
- [34] Baden, J.M. and Cohen, M.N., “Optimal Peak Sidelobe Filters for Biphasic Pulse Compression,” in *Proceedings of the IEEE 1990 International Radar Conference*, pp. 249–252, May 7–10, 1990.
- [35] Griep, K.R., Ritcey, J.A., and Burlingame, J.J., “Design of Mismatched Filters for Pulse Compression in a Multiple User Sonar Ranging System,” in *Proceedings of the 1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1111–1115, November 1–3, 1993.
- [36] Richards, M.A., *Fundamentals of Radar Signal Processing*, McGraw-Hill, New York, 2005.
- [37] Walden, R.H., “Analog-to-Digital Converter Survey and Analysis,” *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, April 1999.
- [38] Le, B., Rondeau, T.W., Reed, J.H., and Bostian, C.W., “Analog-to-Digital Converters,” *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 69–77, November 2005.
- [39] Jonsson, B.E., “A Survey of A/D-Converter Performance Evolution,” in *Proceedings of the 17th IEEE International Conference on Electronics, Circuits, and Systems*, pp. 766–769, December 12–15, 2010.
- [40] Carrara, W.G., Goodman, R.S., and Majewski, R.M., *Spotlight Synthetic Aperture Radar, Signal Processing Algorithm*, Artech House, Boston, 1995.
- [41] Ludeman, L.C., *Fundamentals for Digital Signal Processing*, Harper & Row, New York, 1986.
- [42] Kellogg, W.C., “Digital Processing Rescues Hardware Phase Errors,” *Microwaves and RF*, vol. 21, no. 12, pp. 63–64, 66–67, 80, November 1982.
- [43] Jedwab, J., “A Survey of the Merit Factor Problem for Binary Sequences,” pp. 30–55 in *Lecture Notes in Computer Science vol. 3486, Sequences and their Applications — SETA 2004*, Ed. T. Hellesteth et al., Springer-Verlag, Berlin, 2005.
- [44] Borwein, P., Ferguson, R., and Knauer, J., “The Merit Factor Problem,” pp. 52–70 in *London Mathematical Society Lecture Note Series 352*, Ed. J. McKee and C. Smyth, Cambridge University Press, Cambridge, UK, 2008.
- [45] Brogan, L.W., *Modern Control Theory*, 2d ed., Prentice Hall, Englewood Cliffs, NJ, 1985.

- [46] Nunn, C., "Constrained Optimization Applied to Pulse Compression Codes and Filters," in *Proceedings of the 2005 IEEE International Radar Conference*, pp. 190–194, May 9–12, 2005.
- [47] Levanon, N., "Cross-Correlation of Long Binary Signals with Longer Mismatched Filters," *IEEE Proceedings of Radar, Sonar and Navigation*, vol. 152, no. 6, pp. 372–382, 2005.
- [48] Cohen, M.N., Fox, M.R., and Baden, J.M., "Minimum Peak Sidelobe Pulse Compression Codes," in *Proceedings of the 1990 IEEE International Radar Conference*, Arlington, VA, pp. 633–638, 1990.
- [49] Coxson, G. and Russo, J., "Efficient Exhaustive Search for Optimal-Peak-Sidelobe Binary Codes," *IEEE Trans. AES*, vol. 41, no. 1, pp. 302–308, January 2005.
- [50] Peebles, Jr., P.Z., *Radar Principles*, John Wiley & Sons, New York, 1998.
- [51] Levanon, N. and Mozeson, E., *Radar Signals*, John Wiley & Sons, Hoboken, NJ, 2004.
- [52] Key, E.L., Fowle, E.N., and Haggarty, R.D., "A method of side-lobe suppression in phase-coded pulse compression systems," Technical Report no. 209, M.I.T. Lincoln Laboratory, Lexington, August 1959.
- [53] Ackroyd, M.H. and Ghani, F., "Optimum Mismatched Filters for Sidelobe Suppression," *IEEE Trans. AES*, vol. AES-9, no. 2, pp. 214–218, March 1973.
- [54] Blunt, S.D., and Gerlach, K., "Adaptive Pulse Compression via MMSE Estimation," *IEEE Trans. AES*, vol. 42, no. 2, pp. 572–584, April 2006.

2.10 | PROBLEMS

1. A synthetic aperture radar employs stretch processing. The radar achieves an unweighted, Rayleigh range resolution of 0.2 m. The system is designed to support a 200 m range window, and the pulse width is 10 μsec . What is the filter bandwidth required to support the specified range window?
2. In a system employing stretch processing, the range to the center of the window is 25 km. A target is located within the window at a range of 25.02 km. The LFM waveform has 1.2 GHz bandwidth and a pulse width of 150 μsec . What is the beat frequency associated with the target?
3. A radar system employs a stepped chirp waveform. The waveform consists of 4 pulses. The intrapulse LFM waveform bandwidth is 250 MHz, and the pulse width is 1 μsec . The constant frequency offset is 200 MHz. Assume an amplitude taper is not applied. What is the Rayleigh range resolution achieved by the waveform?
4. A system is designed to reconstruct a stepped chirp waveform by operating on the waveform in the frequency domain. The ADC samples at a 350 MHz rate, and the DFT size is 700. What constraint is placed on the frequency step size Δf ?
5. Using the coefficients in Table 2-2 and equation (107), construct an NLFM waveform with a time-bandwidth product of 50 (e.g., $\beta = 50$ MHz and $\tau = 1$ μsec). What is the peak sidelobe ratio achieved by the waveform? Next, construct an NLFM waveform with a time-bandwidth product of 250 (e.g., $\beta = 250$ MHz and $\tau = 1$ μsec). What is the peak sidelobe ratio achieved by the waveform? Is there an observed relationship between time-bandwidth product and the resultant sidelobe levels? Explain.
6. Using the coefficients in Table 2-2 and equation (107), construct an NLFM waveform with a 200 MHz swept bandwidth and a 2 μsec pulse. Apply a 1/2 cycle of Doppler

shift across the pulse. Generate the compressed response having applied the zero Doppler matched filter. What is the loss in processing gain, the shift in the location of the main lobe peak, and the increase in the peak sidelobe induced by the Doppler shift?

7. A stepped frequency waveform is used to image a 10 m long target. To avoid range ambiguities across the target, what is the largest Δf that may be applied? Using the maximum value of Δf needed to avoid range ambiguities, calculate the number of pulses required to achieve an unweighted, Rayleigh range resolution of 0.1 m.
8. The five-element Barker code is $[1 \ 1 \ 1 \ -1 \ 1]$.
 - a. Convert this to a four-phase sequence using the BTQ transformation.
 - b. What is the maximum element-to-element phase change in the resulting sequence?
9. The 48-element MPS code used in section 2.7.3.2 (and given in Table 4) has a peak sidelobe level of 3. What is the peak sidelobe of the corresponding quadriphase waveform?
10. Compute the optimal ISR filter of length 5 for the five-element Barker code (see problem 8).
11. A filter is needed to suppress the two sidelobes on either side of the main lobe. The other sidelobes are less important. Use the five-element Barker code provided in problem 8 and assume a 5-element filter.
 - a. Define an appropriate weighting matrix for the WLS filter. Use weight values of 1 and 10 along the diagonal.
 - b. Compute the filter and plot the range compressed response.

Optimal and Adaptive MIMO Waveform Design

Joseph R. Guerci

Chapter Outline

3.1	Introduction	87
3.2	Optimum MIMO Waveform Design for the Additive Colored Noise Case	89
3.3	Optimum MIMO Design for Maximizing Signal-to-Clutter Ratio	95
3.4	Optimum MIMO Design for Target Identification	99
3.5	Constrained Optimum MIMO Radar	104
3.6	Adaptive MIMO Radar	109
3.7	Summary	113
3.8	Further Reading	114
3.9	References	114
3.10	Problems	115

3.1 INTRODUCTION

Recent advances in radar front-end hardware such as solid-state transmitters, digital arbitrary waveform generators (DAWGs), active electronically scanned arrays (AESAs), and high-performance embedded computing (HPEC) have afforded an opportunity to re-examine the design of what and how a radar transmits its spatio-temporal radio frequency (RF) signals. Conventional modern radars generally use nonadaptive transmit configurations that typically optimize some metric associated with a multidimensional ambiguity function (e.g., range, Doppler, angle [1]) and do not adapt the transmitter to an ever-changing target and interference channel. Common examples include the very popular constant modulus linear frequency modulated (LFM) waveform and low sidelobe transmit antennas [1]. However, since the output signal-to-interference-plus-noise ratio (SINR) depends on the transmit characteristics, it is natural to ask: for a given channel model, what is the optimum transmit/receive configuration?

Adaptive processing has long been implemented in the receive chain of radar, beginning with automatic gain control and cell-averaging constant false alarm rate (CA-CFAR) [2] all the way to today's space-time adaptive processing (STAP) [3]. However, adaptivity in the transmit chain is virtually nonexistent, save for mode adaptivity such as switching in different nonadaptive waveforms such as pulse repetition frequency (PRF) and bandwidth. This chapter develops the basic theory of optimal transmit/receive design using

a multi-input, multi-output (MIMO) formulation that can account for all potential degrees of freedom (DOFs) such as waveform (fast-time), angle, and polarization. Various applications and examples are provided to further illustrate the potential impact of joint transmit/receive adaptivity.

3.1.1 Organization

This chapter is organized as follows. Section 3.2 introduces the basic MIMO channel formulation and derives the optimal transmitter/receiver configuration for the additive colored noise (ACN) case. Several examples exercising different DOFs (fast-time, spatial) are then presented to illustrate the basic theory and demonstrate the flexibility of the formulation. Next in Section 3.3, the formalism is extended to account for the basic maximizing signal-to-clutter ratio (SCR) problem. In Section 3.4, a basic theory for optimizing transmit/receive configuration for the target ID problem is introduced. The concept of constrained MIMO waveform design is addressed in Section 3.5 to account for important real-world constraints such as constant modulus. Finally, in Section 3.6, the idea of adaptive MIMO waveform design is introduced when the channel must be estimated on the fly.

3.1.2 Key Points

- Fundamental theory for optimum MIMO waveform design
- MIMO waveforms for maximum SINR (additive colored noise case)
- MIMO waveforms for maximizing signal to clutter
- MIMO waveforms for target ID
- Constrained optimum MIMO waveform design
- Adaptive MIMO waveform design

3.1.3 Acronyms

Acronyms that are commonly used in this chapter include the following:

ACN	additive colored noise
AGCN	additive Gaussian colored noise
AOA	angle-of-arrival
CA-CFAR	cell-averaging constant false alarm ratio
CNR	clutter-to-noise ratio
DOFs	degrees-of-freedom
FIR	finite impulse response
GMTI	ground moving target indication
HVT	high value target
i.i.d	independent and identically distributed
LTI	Linear Time Invariant
PRF	pulse repetition frequency
Rx	receiver
SAR	synthetic aperture radar
SCR	signal-to-clutter ratio
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
STAP	space-time adaptive processing
Tx	transmitter
ULA	uniform linear array

3.2 | OPTIMUM MIMO WAVEFORM DESIGN FOR THE ADDITIVE COLORED NOISE CASE

Consider the basic radar block diagram in Figure 3-1. A generally complex-valued and multidimensional transmit signal, $\mathbf{s} \in \mathbb{C}^N$, (i.e., an N -dimensional multi-input (MI) signal), interacts with a target denoted by the target transfer matrix $\mathbf{H}_T \in \mathbb{C}^{M \times N}$. The resulting M -dimensional multi-output (MO) signal (echo), $\mathbf{y} \in \mathbb{C}^M$, is then received along with ACN $\mathbf{n} \in \mathbb{C}^M$. The vector-matrix formulation is completely general inasmuch as any combination of spatial and temporal dimensions can be represented.

For example, the N -dimensional input vector \mathbf{s} could represent the N complex (i.e., in-phase/quadrature, or I/Q [4]) samples of a single-channel transmit waveform $s(t)$, that is,

$$\mathbf{s} = \begin{bmatrix} s(\tau_1) \\ s(\tau_2) \\ \vdots \\ s(\tau_N) \end{bmatrix} \quad (3.1)$$

The corresponding target transfer matrix, \mathbf{H}_T , would thus contain the corresponding samples of the complex target impulse response, $h_T(t)$, which for the causal linear time-invariant (LTI) case would have the form [5]

$$\mathbf{H}_T = \begin{bmatrix} h[0] & 0 & 0 & \cdots & 0 \\ h[1] & h[0] & 0 & \cdots & 0 \\ h[2] & h[1] & h[0] & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ h[N-1] & & & h[1] & h[0] \end{bmatrix} \quad (3.2)$$

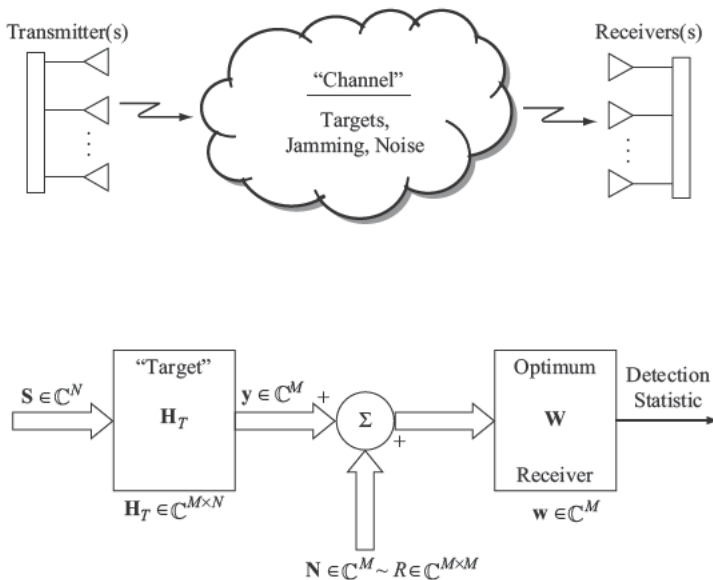


FIGURE 3-1 ■ Fundamental multichannel radar block diagram for the AGCN case. Our objective is to design both the transmit (i.e., \mathbf{s}) and receive (i.e., \mathbf{w}) functions so as to maximize the output SINR given the channel characteristics.

Without loss of generality we have assumed uniform time sampling, that is, $\tau_k = (k-1)T$, where T is a suitably chosen sampling interval [6]. Note also that for convenience and a significant reduction in mathematical nomenclature overhead $N = M$ is used, which is the same number of transmit/receive DOF (e.g., time, space). The reader is encouraged to, where desired, reinstate the inequality and confirm that the underlying equations derived throughout this chapter have the same basic form except for differing vector and matrix dimensionalities. Also note that in general H_T is stochastic.

The formalism is readily extensible to the multiple-transmitter, multiple-receiver case. For example, if there are three independent transmit/receive channels (e.g., an AESA), then the input vector \mathbf{s} of Figure 3-1 would have the form

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \mathbf{s}_3 \end{bmatrix} \in \mathbb{C}^{3N} \quad (3.3)$$

where $\mathbf{s}_i \in \mathbb{C}^N$ denotes the samples (as in (3.1)) of the transmitted waveform out of the i -th transmit channel. The corresponding target transfer matrix would in general have the form

$$H_T = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \in \mathbb{C}^{3N \times 3N} \quad (3.4)$$

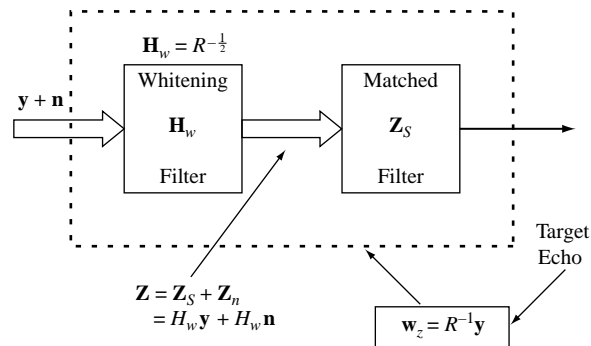
where the submatrix $H_{i,j} \in \mathbb{C}^{N \times N}$ is the transfer matrix between the i -th receive and j -th transmit channels for all time samples of the waveform.

These examples make clear that the matrix–vector, input–output formalism is completely universal and can accommodate whatever transmit/receive DOF desired. Returning to Figure 3-1, we now wish to jointly optimize the transmit/receive functions. We will find it convenient to work backward: to begin by optimizing the receiver as a function of the input and then finally optimizing the input and thus the overall output SINR.

For any finite norm input \mathbf{s} , the receiver that maximizes output SINR for the ACN case is the so-called whitening (or colored noise) matched filter, as shown in Figure 3-2 [7]. Note that for the additive Gaussian colored noise (AGCN) case, this receiver is also statistically optimum [7].

If $R \in \mathbb{C}^{N \times N}$ denotes the total interference covariance matrix associated with \mathbf{n} , which is further assumed to be independent of \mathbf{s} and Hermitian positive definite [8] (guaranteed in practice due to ever-present receiver noise [7]), then the corresponding whitening filter

FIGURE 3-2 ■ The optimum receiver for the ACN case consists of a whitening filter followed by a white noise matched filter.



is given by [7]:

$$H_w = R^{-\frac{1}{2}} \quad (3.5)$$

The reader should verify the whitening properties of (3.5) (see problem 2 and [9]).

The output of the linear whitening filter, $\mathbf{z} \in \mathbb{C}^N$, will consist of signal and noise components, $\mathbf{z}_s, \mathbf{z}_n$, respectively, given by

$$\begin{aligned} \mathbf{z} &= \mathbf{z}_s + \mathbf{z}_n \\ &= H_w \mathbf{y}_s + H_w \mathbf{n} \\ &= H_w H_T \mathbf{s} + H_w \mathbf{n} \end{aligned} \quad (3.6)$$

where $\mathbf{y}_s \in \mathbb{C}^N$ denotes the target echo as shown in Figure 3-2 (i.e., the output of H_T).

Since the noise has been whitened via a linear—in this case full-rank—transformation [7]), the final receiver stage consists of a white noise matched filter of the form (to within a multiplicative scalar)

$$\mathbf{w}_z = \mathbf{z}_s \in \mathbb{C}^N \quad (3.7)$$

The corresponding output SNR is thus given by

$$\begin{aligned} \text{SNR}_o &= \frac{|\mathbf{w}'_z \mathbf{z}_s|^2}{\text{var}(\mathbf{w}'_z \mathbf{z}_n)} \\ &= \frac{|\mathbf{z}'_s \mathbf{z}_s|^2}{\text{var}(\mathbf{z}'_s \mathbf{z}_n)} \\ &= \frac{|\mathbf{z}'_s \mathbf{z}_s|^2}{E\{\mathbf{z}'_s \mathbf{z}_n \mathbf{z}'_n \mathbf{z}_s\}} \\ &= \frac{|\mathbf{z}'_s \mathbf{z}_s|^2}{\mathbf{z}'_s E\{\mathbf{z}_n \mathbf{z}'_n\} \mathbf{z}_s} \\ &= \frac{|\mathbf{z}'_s \mathbf{z}_s|^2}{\mathbf{z}'_s \mathbf{z}_s} \\ &= |\mathbf{z}'_s \mathbf{z}_s| \end{aligned} \quad (3.8)$$

where $\text{var}(\cdot)$ denotes the variance. Note that due to the whitening operation $E\{\mathbf{z}_n \mathbf{z}'_n\} = I$.

In words, the output SNR is proportional to the energy in the whitened target echo. This fact is key to optimizing the input function: Chose \mathbf{s} (the input) to maximize the energy in the whitened target echo:

$$\max_{\{\mathbf{s}\}} |\mathbf{z}'_s \mathbf{z}_s| \quad (3.9)$$

Substituting $\mathbf{z}_s = H_w H_T \mathbf{s}$ into (3.9) yields the objective function that explicitly depends on the input

$$\max_{\{\mathbf{s}\}} |\mathbf{s}' (H' H) \mathbf{s}| \quad (3.10)$$

where

$$H \triangleq H_w H_T \quad (3.11)$$

Recognizing that (3.10) involves the magnitude of the inner product of two vectors \mathbf{s} and $(H'H)\mathbf{s}$, we readily have from the Cauchy–Schwarz theorem [10] the condition that \mathbf{s} must satisfy to yield a maximum, namely, \mathbf{s} must be collinear with $(H'H)\mathbf{s}$:

$$(H'H)\mathbf{s}_{opt} = \lambda_{max}\mathbf{s}_{opt} \quad (3.12)$$

In other words, the optimum input \mathbf{s}_{opt} must be an eigenvector of $(H'H)$ with associated maximum eigenvalue.

The previous set of input–output design equations represents the absolute optimum that any combination of transmit/receive operations can achieve and thus are fundamentally important in ascertaining the value of advanced adaptive methods (e.g., adaptive waveforms, transmit/receive beamforming). Note also that (3.12) can be generalized to the case where the target response is random:

$$E\{H'H\}\mathbf{s}_{opt} = \lambda_{max}\mathbf{s}_{opt} \quad (3.13)$$

In this case, \mathbf{s}_{opt} maximizes the expected value of the output SINR.

Next we illustrate the application of the previously given optimum design equations to the additive colored noise problem arising from a broadband multipath interference source.

EXAMPLE 3.1

Additive Colored Noise Example Arising from Broadband Multipath Interference

This example illustrates the optimum transmit/receive configuration for maximizing output SINR in the presence of colored noise interference arising from a multipath broadband noise source. More specifically, for the single transmit/receive channel case, it derives the optimum transmit pulse modulation (i.e., pulse shape).

Figure 3-3 illustrates a nominally broadband white noise source undergoing a series of multipath scatterings that in turn colors the noise spectrum [11]. Assuming (for simplicity) that the multipath reflections are dominated by several discrete specular reflections, the resultant signal can be viewed as the output of a causal tapped delay line filter (i.e., an FIR filter [5]) of the form

$$h_{mp}[k] = \alpha_0\delta[k] + \alpha_1\delta[k-1] + \cdots + \alpha_{q-1}\delta[k-q-1] \quad (3.14)$$

that is driven by white noise. The corresponding input–output transfer $H_{mp} \in \mathbb{C}^{N \times N}$ is thus given by

$$H_{mp} = \begin{bmatrix} h_{mp}[0] & 0 & \cdots & 0 \\ h_{mp}[1] & h_{mp}[0] & & \vdots \\ \vdots & & \ddots & 0 \\ h_{mp}[N-1] & \cdots & h_{mp}[1] & h_{mp}[0] \end{bmatrix} \quad (3.15)$$

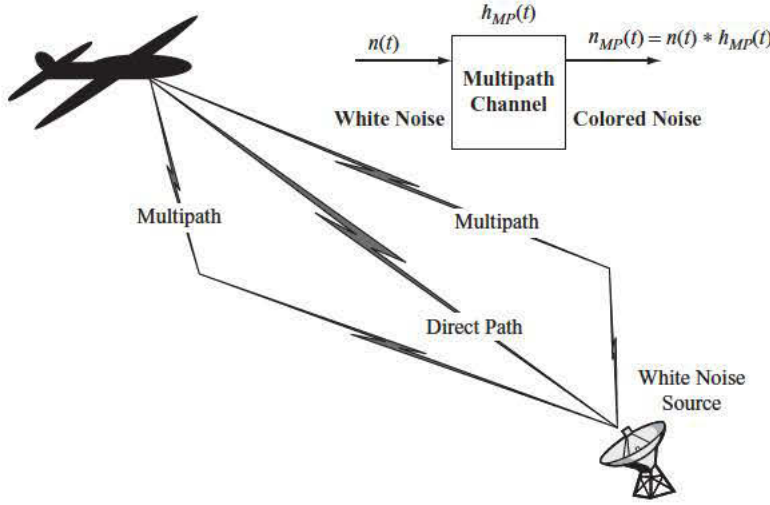


FIGURE 3-3 ■ Illustration of colored noise interference resulting from a broadband (i.e., white noise) source undergoing multipath reflections.

In terms of the multipath transfer matrix, H_{mp} , the colored noise interference covariance matrix is given by

$$\begin{aligned}
 E\{\mathbf{nn}'\} &= E\{H_{mp}\mathbf{vv}'H_{mp}'\} \\
 &= H_{mp}E\{\mathbf{vv}'\}H_{mp}' \\
 &= H_{mp}H_{mp}' \\
 &= R
 \end{aligned} \tag{3.16}$$

where the driving white noise source $\mathbf{v} \in \mathbb{C}^N$ is a zero mean complex vector random variable with an identity covariance matrix:

$$E\{\mathbf{vv}'\} = I \tag{3.17}$$

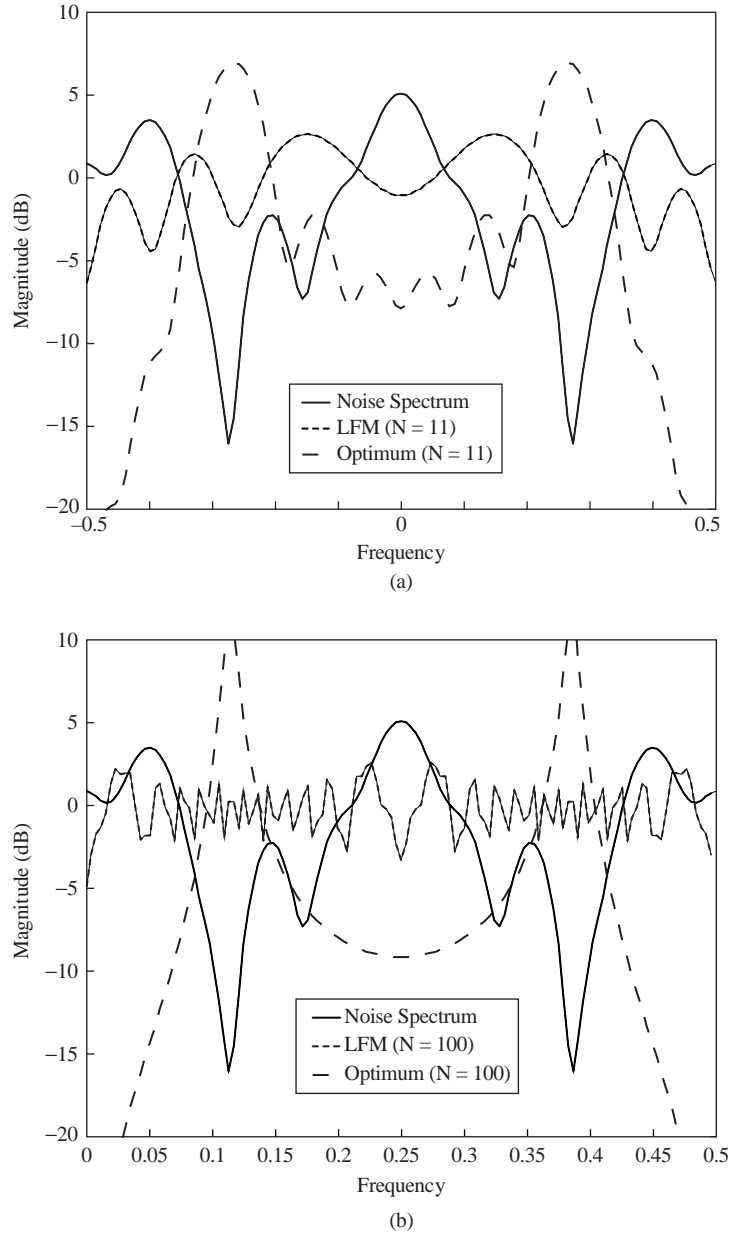
Assuming a unity gain point target at the origin, that is, $h_T[k] = \delta[k]$, yields a target transfer matrix $H_T \in \mathbb{C}^{N \times N}$ given by

$$\begin{aligned}
 H_T &= \begin{bmatrix} h_T[0] & 0 & \cdots & 0 \\ h_T[1] & h_T[0] & & \vdots \\ \vdots & & \ddots & 0 \\ h_T[N-1] & \cdots & h_T[1] & h_T[0] \end{bmatrix} \\
 &= I
 \end{aligned} \tag{3.18}$$

While certainly a more complex (and thus realistic) target model could be assumed, we wish to focus on the impact the colored noise has on shaping the optimum transmit pulse. We will introduce more complex target response models in the target ID section.

Figure 3-4 shows the in-band interference spectrum for the case when $\alpha_0 = 1$, $\alpha_2 = 0.9$, $\alpha_5 = 0.5$, $\alpha_{10} = 0.2$, and all other coefficients are set to zero. The total number of fast-time (range bin) samples was set to both a short-pulse case of $N = 11$ (Figure 3-4a) and a long-pulse case of $N = 100$ (Figure 3-4b). Note that the multipath colors the otherwise flat noise spectrum. Also displayed is the spectrum of a conventional (and thus nonoptimized) LFM pulse with a time-bandwidth product, $\beta\tau$, of 5 (Figure 3-4a) and 50 (Figure 3-4b), respectively [12, 13].

FIGURE 3-4 ■ Spectra of the colored noise interference along with conventional and optimal pulse modulations. (a) Short-pulse case where total duration for the LFM and optimum pulse are set to 11 range bins (fast-time taps). (b) Long-pulse case where total duration for the LFM and optimum pulse are set to 100 range bins. Note that in both cases the optimum pulse attempts to anti-match to the colored noise spectrum under the frequency resolution constraint set by the total pulse width.



Given R from (3.16), the corresponding whitening filter H_w is given by

$$H_w = R^{-\frac{1}{2}} \quad (3.19)$$

Combining (3.19) with (3.18), the total composite channel transfer matrix H is thus given by

$$H = H_w H_T = H_w = R^{-\frac{1}{2}} \quad (3.20)$$

Substituting (3.20) into (3.12) yields

$$R^{-1} \mathbf{s}_{opt} = \lambda \mathbf{s}_{opt} \quad (3.21)$$

That is, the optimum transmit waveform is the maximum eigenfunction associated with the inverse of the interference covariance matrix. The reader should verify that this is also the minimum eigenfunction of the original covariance matrix R and thus can be computed without matrix inversion.

Displayed in Figures 3-4a and 3-4b are the spectra of the optimum transmit pulses obtained by solving (3.21) for the maximum eigenfunction–eigenvalue pair for the aforementioned short- and long-pulse cases, respectively. Note how the optimum transmit spectrum naturally emphasizes portions of the spectrum where the interference is weak—which is an intuitively satisfying result.

The SINR gain of the optimum short pulse, SINR_{opt} , relative to that of a nonoptimized chirp pulse, SINR_{LFM} , is

$$\text{SINR}_{gain} \triangleq \frac{\text{SINR}_{opt}}{\text{SINR}_{LFM}} = 7.0 \text{ dB} \quad (3.22)$$

while for the long-pulse case

$$\text{SINR}_{gain} \triangleq \frac{\text{SINR}_{opt}}{\text{SINR}_{LFM}} = 24.1 \text{ dB} \quad (3.23)$$

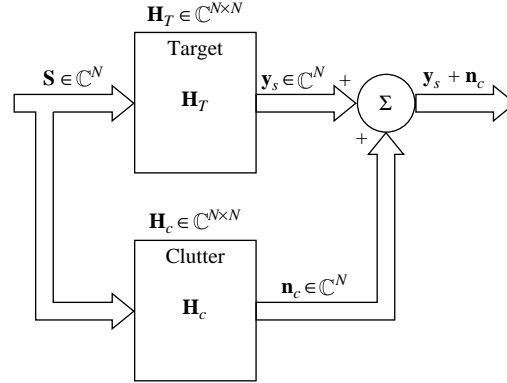
The increase in SINR for the long-pulse case is to be expected since it has finer spectral resolution and can therefore more precisely shape the transmit modulation to antismatch the interference. Of course, the unconstrained optimum pulse has certain practical deficiencies (e.g., poorer resolution, compression sidelobes) compared with a conventional pulse. We will revisit these issues in Section 3.5 where constrained optimization is introduced.

Example 3.1 is similar in spirit to the spectral notching waveform design problem that arises when strong co-channel narrowband interferers are present [14]. In this case it is desirable not only to filter out the interference on receive but also to choose a transmit waveform that minimizes energy in the co-channel bands. The reader is encouraged to experiment with different notched spectra and pulse length assumptions and to apply (3.12) as in example 3.1. Non-impulsive target models can also be readily incorporated.

3.3 | OPTIMUM MIMO DESIGN FOR MAXIMIZING SIGNAL-TO-CLUTTER RATIO

The joint MIMO optimization of the transmit and receive functions for the general additive colored noise plus clutter (signal-dependent noise) has been shown to result in a highly nonlinear problem [15] (though efficient iterative methods have been developed to solve these equations [15]). In practice, however, there is often a natural “separation principle” between additive colored noise (signal independent) and clutter (signal dependent). For example, narrowband electromagnetic interference (EMI) resulting from co-channel interference might require fast-time receiver and transmit spectral notching [14], leaving the slow-time or spatial DOF available for clutter suppression. Similarly, adaptive beamforming for broadband jammer nulling can be separated from the clutter suppression problem in a two-stage approach (see, e.g., [16]). We will thus concentrate in this section on the clutter dominant case and focus solely on maximizing the output signal-to-clutter ratio (SCR).

FIGURE 3-5 ■ Radar signal block diagram for the clutter dominant case illustrating the direct dependency of the clutter signal on the transmitted signal.



Unlike the previous colored noise case in Section 3.2, clutter (i.e., channel reverberations) is a form of signal-dependent noise [17, 18] since the clutter returns depend on the transmit signal characteristics (e.g., transmit antenna pattern and strength, operating frequencies, bandwidths, polarization). Referring to Figure 3-5, the corresponding SCR at the input to the receiver is given by

$$\begin{aligned} \text{SCR} &= \frac{E\{\mathbf{y}'_T \mathbf{y}_T\}}{E\{\mathbf{y}'_c \mathbf{y}_c\}} \\ &= \frac{\mathbf{s}' E\{H'_T H_T\} \mathbf{s}}{\mathbf{s}' E\{H'_c H_c\} \mathbf{s}} \end{aligned} \quad (3.24)$$

where $H_c \in \mathbb{C}^{N \times N}$ denotes the clutter transfer matrix, which is generally taken to be stochastic. Equation (3.24) is a generalized Rayleigh quotient [8] that is maximized when \mathbf{s} is a solution to the generalized eigenvalue problem

$$E\{H'_T H_T\} \mathbf{s} = \lambda E\{H'_c H_c\} \mathbf{s} \quad (3.25)$$

with corresponding maximum eigenvalue. When $E\{H'_c H_c\}$ is positive definite, (3.25) can be converted to an ordinary eigenvalue problem of the form we have already encountered, specifically,

$$E\{H'_c H_c\}^{-1} E\{H'_T H_T\} \mathbf{s} = \lambda \mathbf{s} \quad (3.26)$$

Applying equations (3.25) and (3.26) to the full-up space-time clutter suppression of ground moving target indicator (GMTI) clutter is available in [19]. Due to space limitations, we will instead consider its application to the sidelobe target suppression problem, which is very closely related to the ground clutter interference issue.

EXAMPLE 3.2

Sidelobe Target Suppression

Consider a narrowband $N = 16$ element uniform linear array (ULA) with half-wavelength interelement spacing and a quiescent pattern (Figure 3-6). In addition to the desired target at

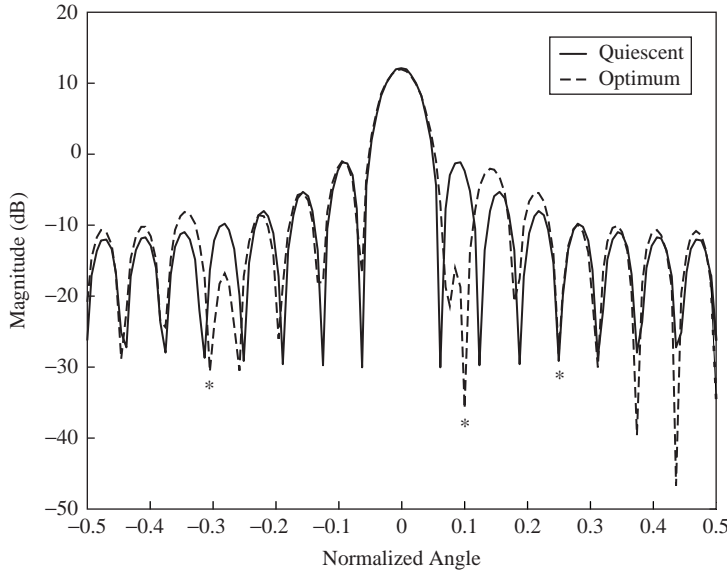


FIGURE 3-6 ■ Illustration of proactive sidelobe target blanking on transmit achieved by maximizing the SCR. Note the presence of nulls in the directions of competing targets while preserving the desired mainbeam response.

a normalized angle of $\bar{\theta} = 0$, there are strong sidelobe targets at $\bar{\theta}_1 = -0.3$, $\bar{\theta}_2 = +0.1$, $\bar{\theta}_3 = +0.25$, where a normalized angle is defined as

$$\bar{\theta} \triangleq \frac{d}{\lambda} \sin \theta \quad (3.27)$$

In (3.27) d is the interelement spacing of the ULA, and λ is the operating wavelength (consistent units and narrowband operation assumed).

The presence of these targets (possibly large clutter discretely) could have been previously detected, thus making their AOAs known. Also, their strong sidelobes could potentially mask weaker mainlobe targets. With this knowledge, it is desired to minimize any energy from these targets leaking into the mainbeam detection of the target of interest by nulling on transmit, or placing transmit antenna pattern nulls in the directions of the unwanted targets.

For the case at hand, the (m, n) -th elements of the target and interferer transfer matrices are given, respectively, by

$$[H_T]_{m,n} = e^{j\varphi} \text{ (const.)} \quad (3.28)$$

$$[H_c]_{m,n} = \alpha_1 e^{j2\pi(m-n)\bar{\theta}_1} + \alpha_2 e^{j2\pi(m-n)\bar{\theta}_2} + \alpha_3 e^{j2\pi(m-n)\bar{\theta}_3} \quad (3.29)$$

where φ is an overall bulk delay (two way propagation) that does not affect the solution to (3.25) and will thus be subsequently ignored, and $[H_c]_{m,n}$ is the (m, n) -th element of the clutter transfer matrix and consists of the linear superposition of the three target returns resulting from transmitting a narrowband signal from the n -th transmit element and receiving it on the m -th receive element of a ULA that uses the same array for transmit and receive [3, 13]. Note that in practice there would be a random relative phase between the signals in (3.29), which for convenience we have ignored but which can easily be accommodated by taking the expected value of the kernel $H_c^H H_c$.

Solving (3.25) for the optimum eigenvector yields the transmit pattern that maximizes the SCR, which is the pattern also displayed in Figure 3-6. The competing target amplitudes were set to 40 dB relative to the desired target and 0 dB of diagonal loading was added to $H_c^H H_c$ to improve numerical conditioning and allow for its inversion. Although this is somewhat

arbitrary, it does provide a mechanism for controlling null depth, that in practice is limited by the amount of transmit channel mismatch [20]. Note the presence of transmit antenna pattern nulls in the directions of the competing targets as desired.

EXAMPLE 3.3

Optimal Pulse Shape for Maximizing SCR

In this simple example, we rigorously verify an intuitively obvious result regarding pulse shape and detecting a point target in uniform clutter: the best waveform for detecting a point target in independent and identically distributed (i.i.d) clutter is itself an impulse (i.e., a waveform with maximal resolution), a well-known result rigorously proven by Manasse [21] using a different method.

Consider a unity point target, arbitrarily chosen to be at the temporal origin. Its corresponding impulse response and transfer matrix are respectively given by

$$h_T[n] = \delta[n] \quad (3.30)$$

and

$$H_T = I_{N \times N} \quad (3.31)$$

where $I_{N \times N}$ denotes the $N \times N$ identity matrix. For uniformly distributed clutter, the corresponding impulse response is of the form

$$h_c[n] = \sum_{k=0}^{N-1} \tilde{\gamma}_k \delta[n - k] \quad (3.32)$$

where $\tilde{\gamma}_i$ denotes the complex reflectivity random variable of the clutter contained in the i -th range cell (i.e., fast-time tap). The corresponding transfer matrix is given by

$$\tilde{H}_c = \begin{bmatrix} \tilde{\gamma}_0 & 0 & 0 & \cdots & 0 \\ \tilde{\gamma}_1 & \tilde{\gamma}_0 & & & \\ \tilde{\gamma}_2 & \tilde{\gamma}_1 & \tilde{\gamma}_0 & & \\ \vdots & & & \ddots & \\ \tilde{\gamma}_{N-1} & \tilde{\gamma}_{N-2} & \tilde{\gamma}_{N-3} & \cdots & \tilde{\gamma}_0 \end{bmatrix} \quad (3.33)$$

Assuming that the $\tilde{\gamma}_i$ values are i.i.d., we have

$$E\{\tilde{\gamma}_i^* \tilde{\gamma}_j\} = P_c \delta[i - j] \quad (3.34)$$

and thus

$$E\left\{[\tilde{H}_c' \tilde{H}_c]_{i,j}\right\} = \begin{cases} 0, & i \neq j \\ (N + 1 - i)P_c, & i = j \end{cases} \quad (3.35)$$

where $[\cdot]_{i,j}$ denotes the (i, j) -th element of the transfer matrix. Note that (3.35) is also diagonal (and thus invertible), but with nonequal diagonal elements.

Finally, substituting (3.31) and (3.35) into (3.26) yields

$$E\{\tilde{H}_c' \tilde{H}_c\}^{-1} \mathbf{s} = \lambda \mathbf{s} \quad (3.36)$$

where

$$E\{\tilde{H}_c' \tilde{H}_c\}^{-1} = \frac{1}{P_c} \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \\ & & \ddots & \\ 0 & & \cdots & d_N \end{bmatrix} \quad (3.37)$$

and

$$d_i \triangleq (N + i - 1)^{-1} \quad (3.38)$$

It is readily verified that the solution to (3.36) yielding the maximum eigenvalue is given by

$$\mathbf{s} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3.39)$$

Thus the optimum pulse shape for detecting a point target is itself an impulse. This should be immediately obvious since it is the shape that excites the range bin only with the target and zeros out all other range bin returns that contain competing clutter.

Of course, transmitting a short pulse (much less an impulse) is problematic in the real world (e.g., creating extremely high peak power pulses) thus an approximation to a short pulse in the form of a spread spectrum waveform (e.g., LFM) is often employed [12]. This example also illuminates that in uniform random clutter nothing is gained by sophisticated pulse shaping for a point target other than to maximize bandwidth (i.e., range resolution) [21]. The interested reader is referred to [19] for further examples of optimizing other DOF (e.g., angle-Doppler) for the clutter mitigation problem.

Up to this point we have been focused on judiciously choosing the transmit/receive DOF to maximize SINR or SCR. In the next section we will extend this framework to the target identification problem.

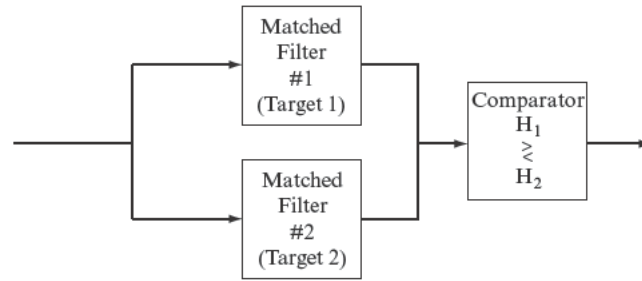
3.4 | OPTIMUM MIMO DESIGN FOR TARGET IDENTIFICATION

Consider the problem of determining target type when two possibilities exist (the multitarget case is addressed later in this section). This can be cast as a classical binary hypothesis testing problem [7]:

$$\begin{aligned} \text{(Target 1)} \quad H_1 : \quad \mathbf{y}_1 + \mathbf{n} &= H_{T_1} \mathbf{s} + \mathbf{n} \\ \text{(Target 2)} \quad H_2 : \quad \mathbf{y}_2 + \mathbf{n} &= H_{T_2} \mathbf{s} + \mathbf{n} \end{aligned} \quad (3.40)$$

where H_{T_1} , H_{T_2} denote the target transfer matrices for targets 1 and 2, respectively. For the AGCN case, the well-known optimum receiver decision structure consists of a bank of matched filters, each tuned to a different target assumption, followed by comparator as shown in Figure 3-7 [7]. Note that (3.40) presupposes that either Target 1 or 2 is present, but

FIGURE 3-7 ■
Optimal receiver
structure for the
binary (two-target)
hypothesis testing
AGCN problem.



not both. Also, it has been tacitly assumed that a binary detection test has been conducted to ensure that a target is indeed present [7]. Alternatively, the null hypothesis (no target present) can be included in the test as a separate hypothesis.

Figure 3-8 illustrates the situation at hand. If Target-1 is present, the observed signal $\mathbf{y}_1 + \mathbf{n}$ will tend to cluster about the #1 point in observation space—which could include any number of dimensions relevant to the target ID problem (e.g., fast-time, angle, Doppler, polarization). The uncertainty sphere (generally ellipsoid for AGCN case) surrounding #1 in Figure 3-7 represents the 1-sigma probability for the additive noise \mathbf{n} —and similarly for #2. Clearly, if \mathbf{y}_1 and \mathbf{y}_2 are relatively well separated, the probability of correct classification is commensurately high.

Significantly, \mathbf{y}_1 and \mathbf{y}_2 depend on the transmit signal \mathbf{s} , as shown in (3.40). Consequently, it should be possible to select an \mathbf{s} that maximizes the separation between \mathbf{y}_1 and \mathbf{y}_2 , thereby maximizing the probability of correct classification under modest assumptions regarding the conditional probability density functions (PDFs) (e.g., unimodality), that is,

$$\max_{\{\mathbf{s}\}} |\mathbf{d}'\mathbf{d}| \quad (3.41)$$

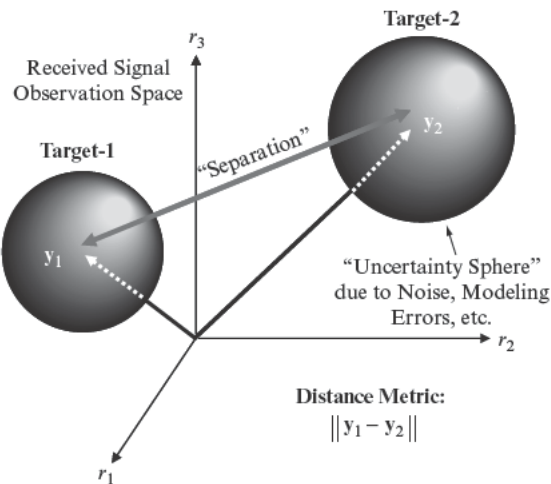


FIGURE 3-8 ■ Illustration of the two-target ID problem. The goal of the joint transmitter/receiver design is to maximally separate the received signals in observation space, which in turn maximizes the probability of correct classification for the additive unimodal monotonic distributed noise case (e.g., AGCN).

where

$$\begin{aligned} \mathbf{d} &\triangleq \mathbf{y}_1 - \mathbf{y}_2 \\ &= H_{T_1} \mathbf{s} - H_{T_2} \mathbf{s} \\ &= (H_{T_1} - H_{T_2}) \mathbf{s} \\ &\triangleq H \mathbf{s} \end{aligned} \quad (3.42)$$

and where

$$H \triangleq H_{T_1} - H_{T_2} \quad (3.43)$$

Substituting (3.42) into (3.41) yields

$$\max_{\{\mathbf{s}\}} |\mathbf{s}' H' H \mathbf{s}| \quad (3.44)$$

This is precisely of the form (3.10) and thus has a solution yielding maximum separation given by

$$(H' H) \mathbf{s}_{opt} = \lambda_{max} \mathbf{s}_{opt} \quad (3.45)$$

It is noted that (3.45) has an interesting interpretation: \mathbf{s}_{opt} is that transmit input that maximally separates the target responses and is thus the maximum eigenfunction of the transfer kernel $H' H$ formed by the difference between the target transfer matrices (i.e., (3.43)). Again if the composite target transfer matrix is stochastic, $H' H$ is replaced with its expected value $E\{H' H\}$ in (3.45).

EXAMPLE 3.4

Two-Target Identification Example

Let $h_1[n]$ and $h_2[n]$ denote the impulse responses of targets #1 and #2, respectively (Figure 3-9). Figure 3-10 shows two different (normalized) transmit waveforms—LFM and optimum (per (3.46))—along with their corresponding normalized separation norms of 0.45 and 1,

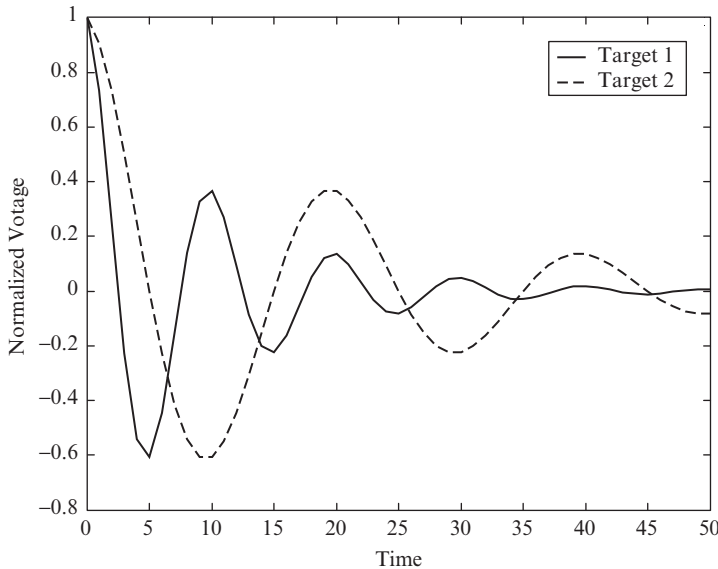
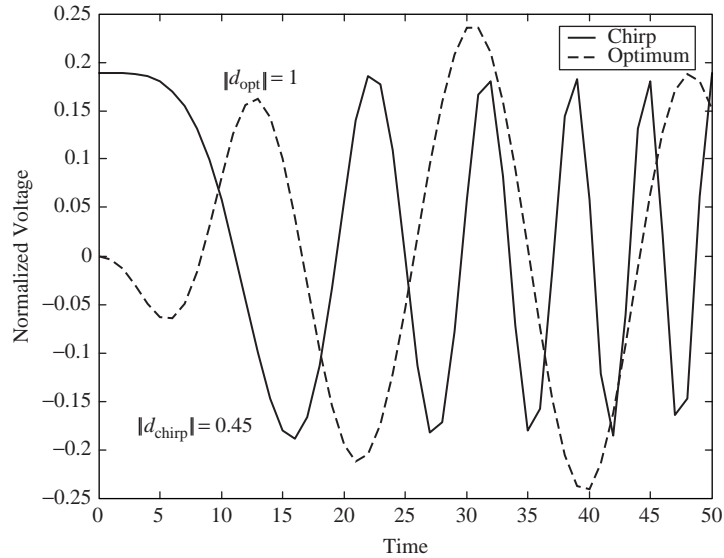


FIGURE 3-9 ■ Target impulse responses utilized for the two-target identification problem.

FIGURE 3-10 ■
Transmit waveforms
employed in the
two-target
identification
example.

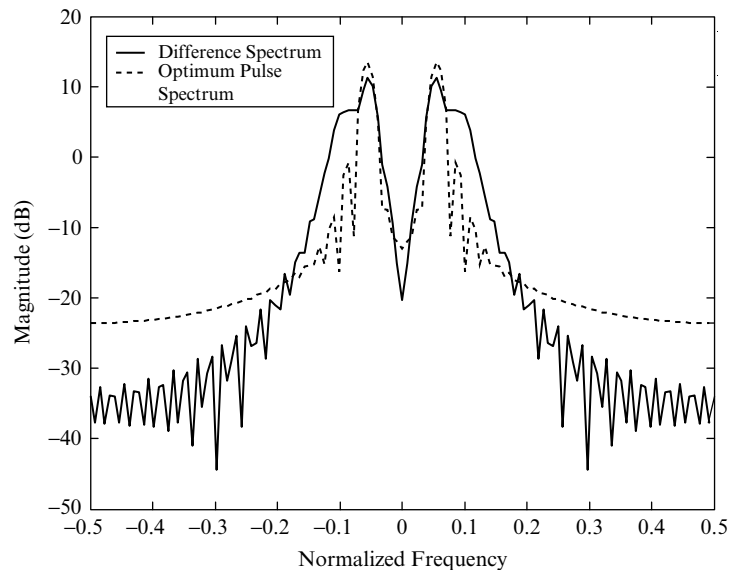


respectively, which corresponds to 6.9 dB improvement in separation. To determine the relative probabilities of correct classification for the different transmit waveforms, one would first need to set the SNR level, which fixes the conditional PDF herein assumed to be circular Gaussian, and then to measure the amount of overlap to calculate the probability [7].

An examination of Figure 3-11 reveals the mechanism by which enhanced separation is achieved. It shows the Fourier spectrum of $H(\omega) = H_{T_1}(\omega) - H_{T_2}(\omega)$, along with that of $S_{\text{opt}}(\omega)$. Note that $S_{\text{opt}}(\omega)$ places more energy in spectral regions where $H(\omega)$ is large (i.e., spectral regions where the difference between targets is large, which is again an intuitively appealing result).

While pulse modulation was used to illustrate the optimum transmit design equations, we could theoretically have used any transmit DOF (e.g., polarization). The choice clearly depends on the application at hand.

FIGURE 3-11 ■
Comparison of the
two-target difference
spectrum and the
optimum pulse
spectrum. Note that
the optimum pulse
emphasizes parts of
the spectrum where
the two targets differ
the most.



Multitarget Case Given L targets in general, we wish to ensure that the L -target response spheres are maximally separated (an inverse sphere packing problem [22]). To accomplish this, we would like to jointly maximize the norms of the set of separations $\{\|\mathbf{d}_{ij}\| \mid i = 1 : L; j = i + 1 : L\}$:

$$\max_{\mathbf{s}} \sum_{i=1}^L \sum_{j=i+1}^L |\mathbf{d}'_{ij} \mathbf{d}_{ij}| \quad (3.46)$$

Since, by definition, \mathbf{d}_{ij} is given by

$$\mathbf{d}_{ij} \triangleq (H_{T_i} - H_{T_j}) \mathbf{s} \triangleq H_{ij} \mathbf{s} \quad (3.47)$$

(3.46) can be rewritten as

$$\max_{\mathbf{s}} \mathbf{s}' \left(\sum_{i=1}^L \sum_{j=i+1}^L H'_{ij} H_{ij} \right) \mathbf{s} \triangleq \mathbf{s}' K \mathbf{s} \quad (3.48)$$

Since $K \in \mathbb{C}^{N \times N}$ is the sum of positive semidefinite matrices, it shares this same property, and thus the optimum transmit input satisfies

$$K \mathbf{s}_{opt} = \lambda_{max} \mathbf{s}_{opt} \quad (3.49)$$

EXAMPLE 3.5

Multitarget Identification

Figure 3-12 depicts the impulse responses of three different targets, two of which are the same as in Example 3.4. Solving (3.48) and (3.49) yields an optimally separating waveform

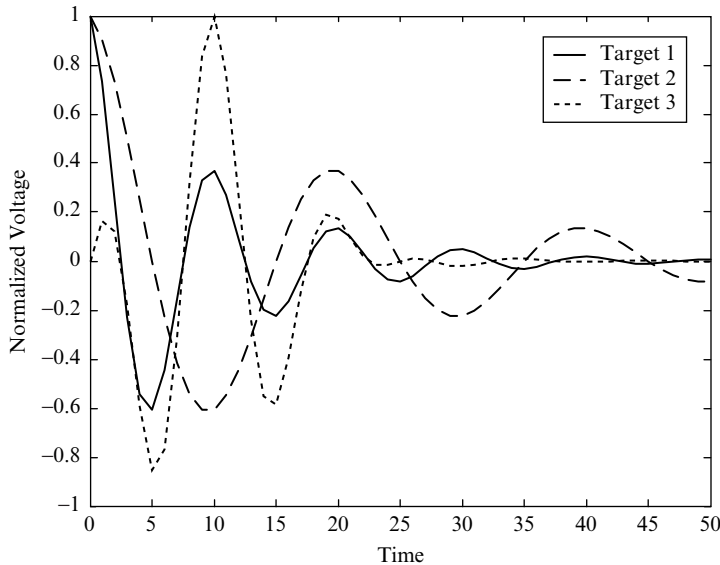


FIGURE 3-12 ■ Target impulse responses used for the three-target identification problem.

whose average separation defined by (3.46) is 1.0. This is compared with 0.47 for the LFM of Example 3.4, an improvement of 6.5 dB, which is slightly less than the previous two-target example. As expected, the optimum waveform significantly outperforms the unoptimized pulse waveform such as the LFM.

3.5 | CONSTRAINED OPTIMUM MIMO RADAR

Often there are a number of practical considerations may preclude transmitting the unconstrained optimum solutions developed so far. We will thus consider two cases of constrained optimization: linear and nonlinear.

Case 3.1: Linear Constraints

Consider the linearly constrained version of the input optimization problem:

$$\max_{\{\mathbf{s}\}} |\mathbf{s}' H' H \mathbf{s}| \quad (3.50)$$

$$\text{subject to: } G \mathbf{s} = \mathbf{0} \quad (3.51)$$

where $G \in \mathbb{C}^{Q \times N}$. To avoid the overly constrained case, it is assumed that $Q < N$. For example, the rows of G could represent steering vectors associated with known interferers such as unwanted targets or clutter discretized to which we wish to apply transmit nulls.

Equation (3.51) defines the feasible solution subspace for the constrained optimization problem. It is straightforward to verify that the projection operator

$$P = I - G' (G G')^{-1} G \quad (3.52)$$

projects any $\mathbf{x} \in \mathbb{C}^N$ into the feasible subspace [23]. Thus, we can first apply the projection operator then perform an unconstrained subspace optimization to obtain the solution to (3.50) and (3.51), that is,

$$\max_{\{\mathbf{s}\}} |\mathbf{s}' P' H' H P \mathbf{s}| \quad (3.53)$$

From (3.53) it is readily apparent that the constrained optimum transmit input satisfies

$$P' H' H P \mathbf{s}_{opt} = \lambda_{max} \mathbf{s}_{opt} \quad (3.54)$$

EXAMPLE 3.6

Prenulling on Transmit

If there are known AOAs for which it is desired not to transmit (e.g., unwanted targets, clutter discrete, keep-out zones), it is possible to formulate a linearly constrained optimization accordingly.

Assume that there is a desired target at $\bar{\theta}_T$ as well as two keep-out angles (normalized) $\bar{\theta}_{I_1}$ and $\bar{\theta}_{I_2}$. The corresponding elements of the target transfer matrix $H_T \in \mathbb{C}^{N \times N}$, assuming an N -element ULA, are thus given by

$$[H_T]_{m,n} = e^{j2\pi(m-n)\bar{\theta}_T} \quad (3.55)$$

where $[H_T]_{m,n}$ denotes the (m, n) -th element of the target transfer matrix.

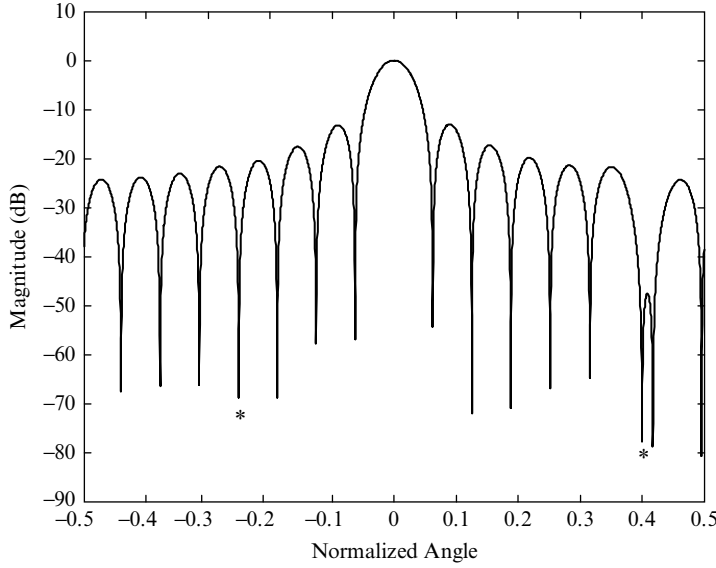


FIGURE 3-13 ■ Example of a linearly constrained optimization in which two interferers are removed via the projection optimization approach.

The keep-out constraints have the form

$$\begin{aligned} \mathbf{0} &= G\mathbf{s} \\ &= \begin{bmatrix} \mathbf{s}'_{I_1} \\ \mathbf{s}'_{I_2} \end{bmatrix} \mathbf{s} \end{aligned} \quad (3.56)$$

where

$$\mathbf{s}_{I_k} = \begin{bmatrix} 1 \\ e^{j2\pi\bar{\theta}_{I_k}} \\ \vdots \\ e^{j2\pi(N-1)\bar{\theta}_{I_k}} \end{bmatrix} \quad (3.57)$$

Figure 3-13 shows the resulting constrained optimum transmit pattern for the case where $\bar{\theta}_T = 0$, $\bar{\theta}_{I_1} = -0.25$, $\bar{\theta}_{I_2} = 0.4$. As expected a peak is placed in the desired target direction with nulls simultaneously placed in the keep-out directions.

Case 3.2: Nonlinear Constraints

In practice other generally nonlinear constraints may arise. One family of such constraints relates to the admissibility of transmit waveforms, such as the class of constant modulus and stepped frequency waveforms [12], to name but a few.

For example, if it is desired to transmit a waveform that is nominally of the LFM type (or any other prescribed type) but that is allowed to modestly deviate to better match the channel characteristics, then the nonlinear constrained optimization has the form

$$\max_{\{\mathbf{s}\}} |\mathbf{s}' H' H \mathbf{s}| \quad (3.58)$$

$$\text{subject to: } \|\mathbf{s} - \mathbf{s}_{LFM}\| \leq \delta \quad (3.59)$$

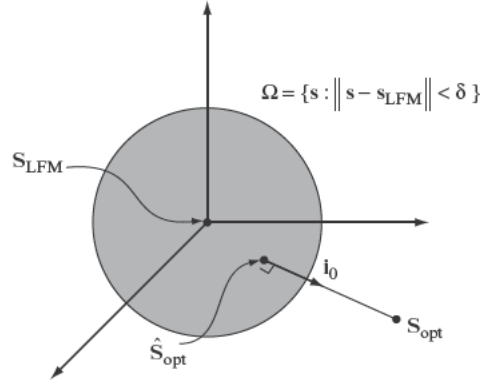


FIGURE 3-14 ■ Illustration of a constrained optimization in which the signal should lie within a subspace (in this case convex) defined to be close to a prescribed transmit input (in this case an LFM waveform). The optimum relaxed projection is the point closest to the unconstrained optimum but still residing in the subspace.

The previous and similar problems cannot generally be solved in closed form. However, approximate methods can yield satisfactory results, and we will consider two here that are based on very different approaches. These simpler methods could form the basis of more complex methods, such as seeding nonlinear search methods.

Relaxed Projection Approach Figure 3-14 depicts the constrained optimization problem in (3.58) and (3.59). It shows the general situation in which the unconstrained optimum solution does not reside within the constrained (i.e., admissible) subspace Ω . In this particular case, the admissible subspace is a convex set [24], defined as

$$\Omega = \{s : \|s - s_{LFM}\| \leq \delta\} \quad (3.60)$$

From Figure 3-14 it is also immediately evident that the admissible waveform closest (in a normed sense) to the unconstrained optimum s_{opt} lies on the surface of Ω along the direction \mathbf{i}_o , which is the unit norm vector that points from s_{LFM} to s_{opt} , i.e.,

$$\mathbf{i}_o \triangleq \frac{s_{opt} - s_{LFM}}{\|s_{opt} - s_{LFM}\|} \quad (3.61)$$

Thus, the constrained waveform that is closest in norm to s_{opt} is given by

$$\hat{s}_{opt} = s_{LFM} + \delta \mathbf{i}_o \quad (3.62)$$

Note that if δ is allowed to relax to the point where $\delta = \|s_{opt} - s_{LFM}\|$, then $\hat{s}_{opt} = s_{opt}$.

EXAMPLE 3.7

Relaxed Projection Example

Here an LFM similarity constraint is imposed on the multipath interference problem considered in Example 3.1. Specifically, in Figure 3-15, we plot the loss in SINR relative to the unconstrained long-pulse optimum solution originally obtained in Example 3.1 as a function

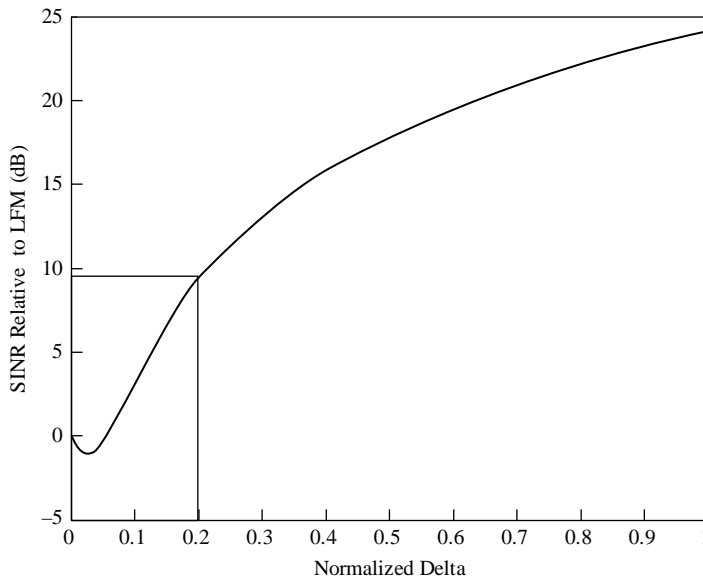


FIGURE 3-15 ■ Illustration of the relaxed projection method for constrained optimization. The plot shows the SINR improvement relative to the unoptimized LFM waveform of example 3.1 versus the normalized relaxation parameter δ . Note that for even a modest relaxation of 20% a nearly 10 dB gain in performance is achieved.

of δ , which is varied between $0 \leq \delta \leq \|\mathbf{s}_{opt} - \mathbf{s}_{LFM}\|$. Note that for this example improvement generally monotonically increases with increasing δ (except for a very small region near the origin) and that sizeable SINR improvements can be achieved for relatively modest values of the relaxation parameter. In other words, a waveform with LFM-like properties can be constructed that still achieves significant SINR performance gains relative to an unoptimized LFM.

Figure 3-16 shows the spectra of the unoptimized LFM of example 3.1 along with the unconstrained optimum and the relaxed projection pulse with a 20% relaxation parameter. Note how the relaxed pulse is significantly closer to the original LFM spectrum yet still achieves nearly a 10 dB improvement in SINR relative to the LFM waveform.

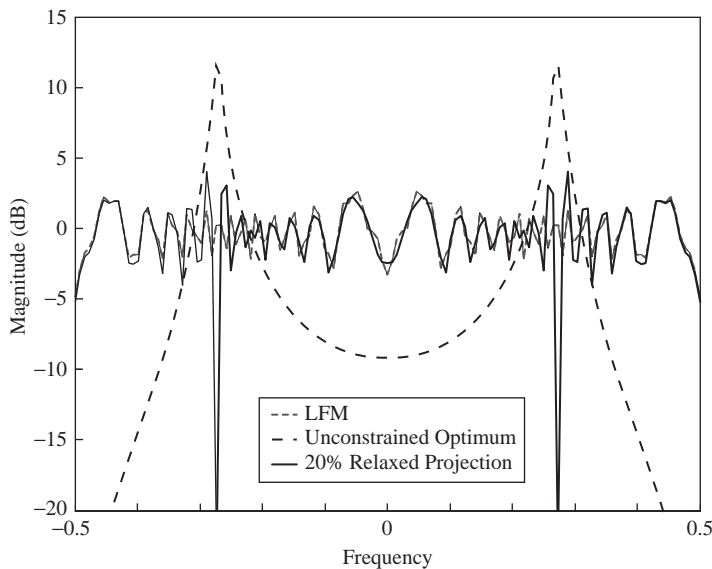


FIGURE 3-16 ■ Comparison of the pulse spectra for the original LFM, unconstrained optimum, and 20% relaxed projection. Note how the relaxed pulse retains LFM-like spectral characteristics (and thus enhanced resolution for example) yet still achieves a 10 dB SINR improvement.

Constant Modulus and the Method of Stationary Phase As has become apparent from the previous examples, spectral shaping plays a key role in achieving matching gains. The stationary phase method has been applied to the problem of creating a nonlinear frequency modulated (NLFM) pulse (and thus constant modulus in the sense that the modulus of the baseband complex envelope is constant, i.e., $|s(t)| = \text{constant}$) with a prescribed magnitude spectrum [5, 12].

Specifically, under fairly general conditions [5, 12] it is possible to relate instantaneous frequency ω of a NLFM waveform to time t [5, 12]:

$$\frac{1}{2\pi} \int_{-\infty}^{\omega} |S(\omega)|^2 d\omega = k \int_0^t dt = kt \quad (3.63)$$

$$t \in [0, T]$$

where $|S(\omega)|$ is the magnitude spectrum of the optimum pulse. Here we have assumed a constant modulus for the NLFM waveform resulting in an integral that is simply proportional to time (see [5, 12] for the more general nonconstant modulus case) as well as a finite and causal pulse.

Solving for ω as a function of t in (3.63) yields the frequency modulation that will result in a transmit pulse with a magnitude spectrum equal to $|S(\omega)|$, to within numerical and other theoretical limitations [5, 12].

EXAMPLE 3.8

NLFM to Achieve Constant Modulus

Here we use the method of stationary phase to design a constant modulus NLFM pulse that matches the magnitude spectrum of the optimum pulse derived for the multipath interference problem considered in Example 3.1.

Figure 3-17 shows the solution to (3.63) (i.e., ω versus t) along with the optimum pulse spectrum from Example 3.1 (long-pulse case). Note that as one would intuit, the frequency modulation dwells at frequencies where peaks in the optimum pulse spectrum occur and conversely note the regions in which the modulation speeds up to avoid frequencies where the optimum pulse spectrum has nulls or lower energy content.

The constant modulus NLFM waveform so constructed was able to achieve an output SINR that was within 6.0 dB of optimum compared with a 24 dB loss using an LFM waveform of same energy and duration.

It is natural to ask if a NLFM waveform with the same spectral magnitude as the optimum pulse (but not necessarily the same phase) will enjoy some (if not all) of the matching gains. For the steady-state case (infinite time duration) this is indeed true, since from Parseval's [5] theorem the output energy is related to only the spectral magnitudes (i.e., without their phases) of the input pulse and channel transfer function, that is,

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 |S(\omega)|^2 d\omega \quad (3.64)$$

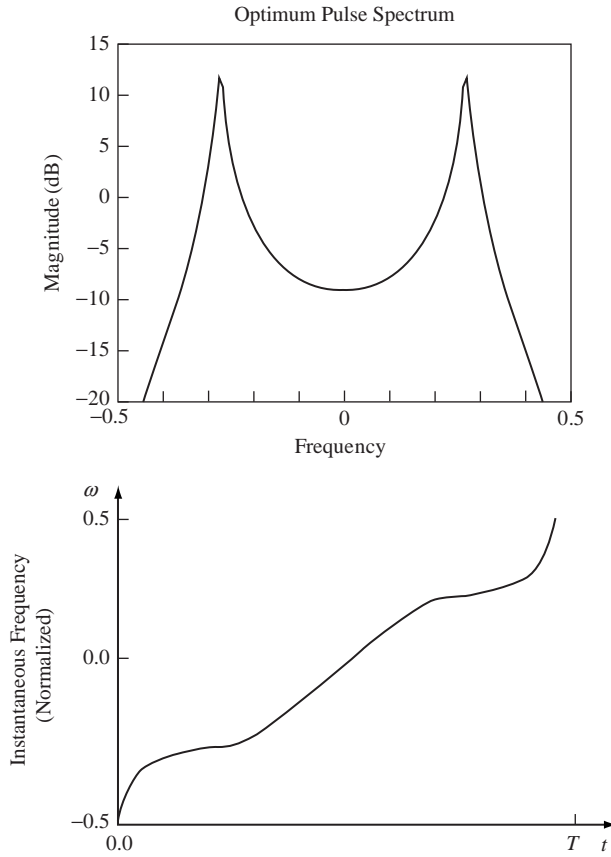


FIGURE 3-17 ■ Illustration of the use of the method of stationary phase to create a constant modulus NLFM pulse whose spectral magnitude matches that of the optimum pulse. The NLFM pulse was able to achieve an output SINR that was within 6.0 dB of the optimum compared with a 24 dB loss using an LFM waveform of same energy and duration.

where $Y(\omega)$, $H(\omega)$, and $S(\omega)$ denote the Fourier transforms of the channel output, channel impulse response, and input pulse, respectively. Note that the output energy in (3.64) depends on the spectral magnitude of the input pulse (steady-state)—not the phase. Thus, in theory an NLFM waveform that exactly matches the optimum pulse magnitude spectrum will achieve the same matching gains in the steady-state limit (infinite pulse duration) for all square integrable (finite norm) functions.

3.6 | ADAPTIVE MIMO RADAR

Section 3.2 derived the optimal multidimensional transmit/receive (i.e., MIMO) design equations that assumed exact knowledge (deterministic or statistical) of the channel (target and interference). However, as those familiar with real-world radar are well aware, channel characterization in large part must be performed on the fly, or adaptively. This is simply a result of the dynamic nature of real-world targets and especially interference.

While a plethora of techniques have been developed for radar receiver adaptivity, estimating requisite channel characteristics for adapting the transmit function—especially for transmit-dependent interference such as clutter—is a relatively new endeavor. In this chapter, we explore several approaches for addressing the adaptive MIMO optimization problem.

In Section 3.6.1, we introduce techniques for the case when the channel characteristics are independent of the transmit input—an example of which is additive colored noise jamming. Perhaps not surprisingly, given the transmit independence, the channel estimation techniques are essentially those often invoked in receive-only adaptivity (e.g., STAP [3]).

Section 3.6.2 introduces adaptive MIMO techniques for dynamic transmit array calibration, including the special case of cohere on target. This latter method enables the cohering of RF transmissions of distributed radars for a specific high-value target (HVT) of interest. The methods using the orthogonality approach to waveform design first introduced by Bliss and Forsythe [25] can thus be viewed as a means for estimating the MIMO channel. However, once an estimate of the channel is made, the optimum MIMO transmit/receive functions should be employed.

3.6.1 Transmit-Independent Channel Estimation

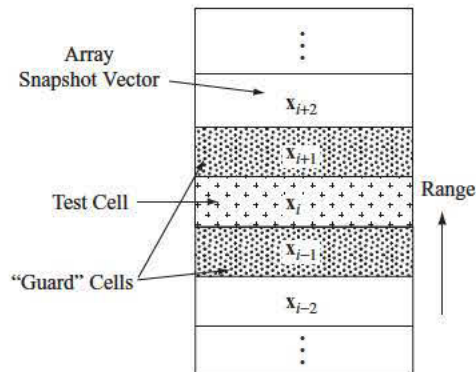
As mentioned previously, a multitude of techniques has been developed for the so-called transmit-independent case. A classic example is additive noise jamming [20]. For the case where no a priori knowledge is available, the baseline method of sample covariance estimation—and its many variants such as diagonal loading and principal components [26, 27]—is often used. In addition to its statistical optimality properties (it is the maximum likelihood solution for the i.i.d. additive Gaussian noise case [7]), efficient parallel processing implementations have been developed facilitating its real-time operation [28].

Figure 3-18 depicts a common procedure for estimating additive, transmit-independent interference statistics. Specifically, the interference covariance matrix, $R \in C^{N \times N}$, is approximated by $\hat{R} \in C^{N \times N}$, where

$$\hat{R} = \frac{1}{L} \sum_{q \in \Omega} \mathbf{x}_q \mathbf{x}_q' \quad (3.65)$$

where $\mathbf{x}_q \in C^N$ denotes the N -dimensional receive array snapshot (e.g., spatial, spatiotemporal) corresponding to the q -th independent temporal sample (e.g., a range or Doppler bin), and L denotes the number of i.i.d. samples selected from a suitable set of training samples Ω to form the summation. As depicted in Figure 3-18, this training region is often chosen to be close in range to the range cell of interest (though there are many variants of this). If, moreover, the selected samples are Gaussian and i.i.d., then (3.65) can be shown to provide the maximum likelihood estimate of R [7]. We illustrate this approach in the following example.

FIGURE 3-18 ■
Illustration of a common method for estimating the interference statistics for the additive transmit-independent case.



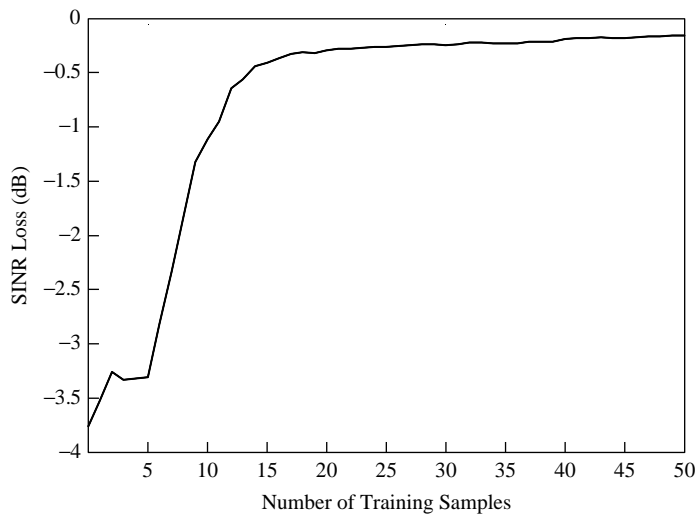


FIGURE 3-19 ■ Effect of sample support on output SINR loss for the multipath interference scenario of example 3.1.

EXAMPLE 3.9

Adaptive Multipath Interference Mitigation

This is a repeat of example 3.1 with the notable exception of unknown interference statistics that must be estimated on the fly. As a consequence, an estimate of the covariance matrix is used in (3.5) for the whitening filter rather than the actual covariance, as was the case in Section 3.2.

Plotted in Figure 3-19 is the overall output SINR loss relative to the optimum for the short-pulse case of example 3.1 as a function of the number of independent samples used in (3.65). The results shown were based on 50 Monte Carlo trials (root mean square average) with a jammer-to-noise ratio of 50 dB and a small amount of diagonal loading to allow for inversion when the number of samples is less than 11 (positive semidefinite case).

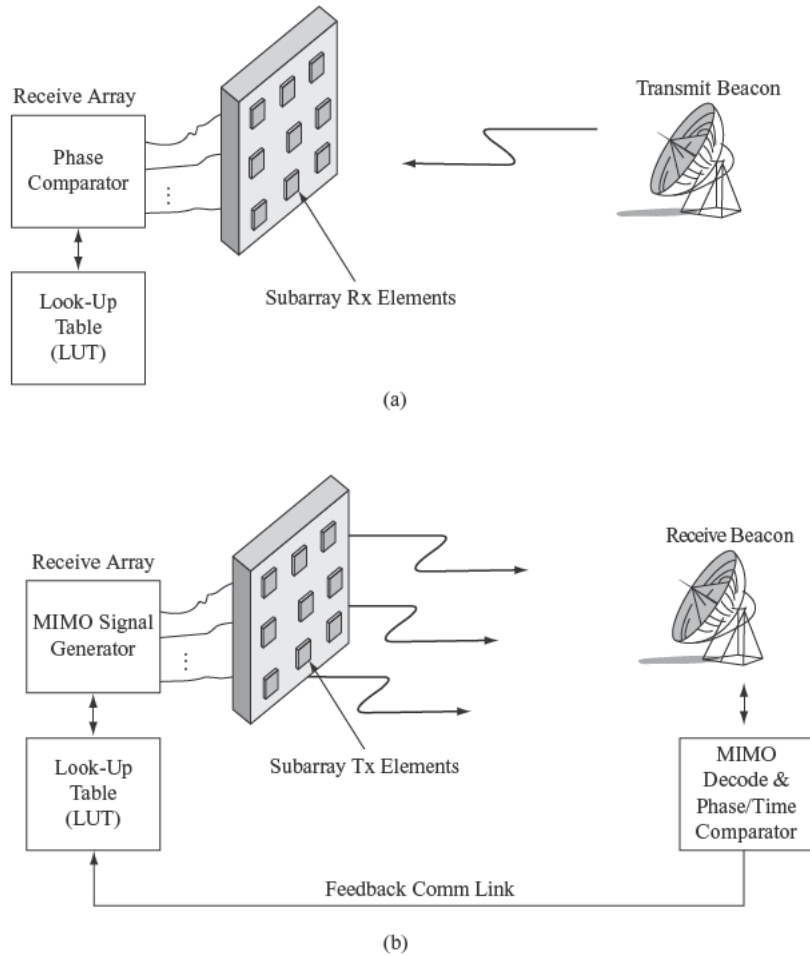
It is interesting to note the rapid convergence and contrast this with SINR loss performance for adaptive beamforming, which is generally significantly slower because we are estimating the single dominant eigenvalue–eigenvector pair. For an authoritative examination of principal components estimation and convergence properties, see [29].

3.6.2 Dynamic MIMO Calibration

Perhaps the earliest MIMO radar techniques have their origins in transmit antenna array calibration [30, 31]. While techniques for estimating the receive array manifold using cooperative or noncooperative sources have existed for quite some time [30], methods for dynamically calibrating the transmit array manifold (e.g., AESAs) in situ are relatively recent developments [31].

Figure 3-20 provides an example of using MIMO techniques to dynamically calibrate an AESA radar. Orthogonal waveforms are simultaneously transmitted from each transmit/receive site of an AESA (typically a single subarray AESA). A cooperative receiver then decodes each individual signal, calculates the relative phases (or time delays), and transmits this information back to the radar. By repeating this process for different orientations, a detailed look-up table for the transmit steering vectors can be constructed

FIGURE 3-20 ■
Illustration of a MIMO-based in situ calibration technique for an AESA radar [30, 31]. (a) Conventional receive array calibration using a known in-band illuminator. (b) The MIMO approach for calibrating the transmit array. (From Steinberg and Yadin [30] and Guerri and Jaska [31]. With permission.)



onboard the radar platform. This in situ approach is basically a necessity for very large AESAs in space since rigidity, which requires mass/weight, is not sufficient to maintain prelaunch calibration [31].

EXAMPLE 3.10

MIMO Cohere-on-Target

An interesting special case of the previously mentioned dynamic in situ calibration procedure is when transmit calibration is performed for a distributed radar focused on a single HVT, as described by Coutts et al. [32].

Consider Figure 3-21, which depicts an airborne HVT that can be detected simultaneously by two geographically disparate radars. Given the HVT nature of the target, it is desired to have the two radars work coherently to maximize the overall SNR at each radar. To achieve on-target coherency, the two waveforms from each radar need to interfere constructively. To accomplish this, however, requires precise knowledge of the transmit pathways to a fraction of a wavelength [32]—essentially a dynamic calibration.

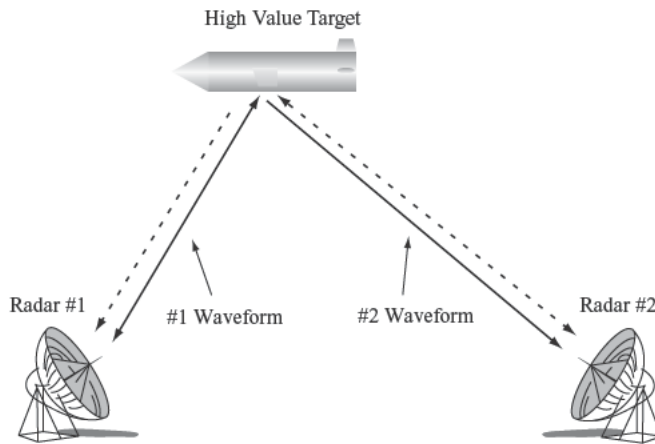


FIGURE 3-21 ■ Illustration of the MIMO cohere-on-target approach for maximizing distributed radar performance. (From Coutts et al. [32]. With permission.)

Drawing on MIMO-based calibration concepts, the requisite relative time delays between the two radars (as seen by the target) can be estimated by simultaneously transmitting orthogonal waveforms, which are then detected and processed in each radar as follows:

- At each radar, the known one-way time delay to the target is subtracted from the total transit time for the sister radar (precise time synchronization is assumed). The remaining time delay is thus due to the first leg of the bistatic path (see Figure 3-21).
- By precompensating a joint waveform in each radar, the two waveforms can be made to cohere on the target—resulting in a 3 dB SNR boost (ideally). If the previous procedure is repeated for N radars, as much as a $10 \log N$ dB gain in SNR is theoretically achievable.

While relatively straightforward to describe, the aforementioned procedure is replete with many real-world difficulties including target motion compensation to a fraction of a wavelength and precise phase and timing stability.

As mentioned previously, the orthogonal waveform MIMO radar approach can provide a means for adaptively estimating the composite target-interference channel since the individual input-output responses can, under certain circumstances, be resolved simultaneously. However, once an estimate of the composite channel is achieved, the optimal MIMO transmit/receive configuration derived in this chapter should be employed to maximize SINR.

3.7 | SUMMARY

In this chapter, the fundamental theory for joint optimization of the transmit and receive functions was developed from first principles and applied to the maximization of SINR, SCR, and correct classification for the target ID problem. Constrained optimization was introduced to address additional requirements that often arise in practice, such as the use of constant modulus waveforms to maximize transmitter efficiency. Lastly, basic adaptive methods were introduced to address the real-world issue of estimating the requisite channel information required when it is not available a priori.

3.8 | FURTHER READING

Further in-depth details and examples on optimum and adaptive MIMO waveform design, including knowledge-aided methods, can be found in [19]. Further details on the orthogonal MIMO approach can be found in [25].

3.9 | REFERENCES

- [1] M. A. Richards, *et al.*, *Principles of Modern Radar, Vol I: Basic Principles*. Raleigh, NC: SciTech, 2010.
- [2] J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals, volume 24 of MIT Radiation Laboratory Series*. New York: McGraw-Hill, 1950.
- [3] J. R. Guerci, *Space-time adaptive processing for radar*. Norwood, MA: Artech House, 2003.
- [4] D. K. Barton, *Modern radar system analysis*. Norwood, MA: Artech House, 1988.
- [5] A. Papoulis, *Signal analysis*: McGraw-Hill New York, 1984.
- [6] A. Papoulis, *Circuits and Systems: A Modern Approach*. New York: Holt, Rinehart and Winston, 1980.
- [7] H. L. V. Trees, *Detection, Estimation and Modulation Theory. Part I*. New York: Wiley, 1968.
- [8] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge [England] ; New York: Cambridge University Press, 1990.
- [9] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*: McGraw-Hill New York, 1991.
- [10] D. A. Pierre, *Optimization Theory With Applications*: Courier Dover Publications, 1986.
- [11] J. R. Guerci and S. U. Pillai, "Theory and application of optimum transmit-receive radar," in *Radar Conference, 2000. The Record of the IEEE 2000 International*, 2000, pp. 705–710.
- [12] C. E. Cook and M. Bernfeld, *Radar signals*: Academic Press New York, 1967.
- [13] M. A. Richards, *Fundamentals of radar signal processing*: McGraw-Hill, 2005.
- [14] M. J. Lindenfeld, "Sparse frequency transmit-and-receive waveform design," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 40, pp. 851–861, 2004.
- [15] S. U. Pillai, *et al.*, "Optimal transmit-receiver design in the presence of signal-dependent interference and channel noise," *Information Theory, IEEE Transactions on*, vol. 46, pp. 577–584, 2000.
- [16] D. J. Rabideau, "Closed Loop Multistage Adaptive Beamforming," *Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems and Computers*, pp. 98–102, 1999.
- [17] H. L. Van Trees and E. Detection, "Modulation Theory, Part II," ed: New York: John Wiley and Sons, 1971.
- [18] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*: Krieger Publishing Co., Inc., 1992.
- [19] J. R. Guerci, *Cognitive Radar: The Knowledge-Aided Fully Adaptive Approach*. Norwood, MA: Artech House, 2010.
- [20] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*: SciTech Publishing, 2003.
- [21] R. Manasse, "The Use of Pulse Coding to Discriminate Against Clutter," *Defense Technical Information Center (DTIC)*, vol. AD0260230, June 7, 1961.

- [22] W. Y. Hsiang, "On the sphere packing problem and the proof of Kepler's conjecture," *Internat. J. Math.*, vol. 4, pp. 739–831, 1993.
- [23] W. Gander, *et al.*, "A constrained eigenvalue problem," *Linear Algebra Appl.*, vol. 114, pp. 815–839, 1989.
- [24] D. C. Youla and H. Webb, "Image Restoration by the Method of Convex Projections: Part 1? Theory," *Medical Imaging, IEEE Transactions on*, vol. 1, pp. 81–94, 1982.
- [25] D. W. Bliss and K. W. Forsythe, "Multiple-input multiple-output (MIMO) radar and imaging: degrees of freedom and resolution," presented at the Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on, 2003.
- [26] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 24, pp. 397–401, 1988.
- [27] A. M. Haimovich and M. Berin, "Eigenanalysis-based space-time adaptive radar: performance analysis," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 33, pp. 1170–1179, 1997.
- [28] A. Farina and L. Timmoneri, "Real-time STAP techniques," *Electronics & Communication Engineering Journal*, vol. 11, pp. 13–22, 1999.
- [29] S. T. Smith, "Covariance, subspace, and intrinsic Cramer-Rao bounds," *Signal Processing, IEEE Transactions on*, vol. 53, pp. 1610–1630, 2005.
- [30] B. D. Steinberg and E. Yadin, "Self-Cohering an Airborne e Radio Camera," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-19, pp. 483–490, 1983.
- [31] J. Guerçi and E. Jaska, "ISAT - innovative space-based-radar antenna technology," in *Phased Array Systems and Technology, 2003. IEEE International Symposium on*, 2003, pp. 45–51.
- [32] S. Coutts, *et al.*, "Distributed Coherent Aperture Measurements for Next Generation BMD Radar," in *Sensor Array and Multichannel Processing, 2006. Fourth IEEE Workshop on*, 2006, pp. 390–393.
- [33] P. G. Grieve and J. R. Guerçi, "Optimum matched illumination-reception radar," ed: US Patent 5,175,552, 1992.
- [34] J. R. Guerçi and P. G. Grieve, "Optimum matched illumination waveform design process," ed: US Patent 5,121,125, 1992.
- [35] A. Farina and F. A. Studer, "Detection with High Resolution Radar: Great Promise, Big Challenge," *Microwave Journal*, May 1991 1991.
- [36] J. R. Guerçi, *et al.*, "Constrained optimum matched illumination-reception radar," ed: US Patent 5,146,229, 1992.

3.10 | PROBLEMS

1. A noncasual impulse response can have nonzero values for negative time indices, that is, $h(-k) \neq 0$ for some positive k value. This can arise when the impulse response is not associated with time, such as the case when k is a spatial index or when time samples are processed in batch (buffered) fashion. Rederive the H matrix of equation (3.2) when the impulse response has values from $-M$ to M .
2. Verify that the whitening filter of equation (3.5) (i.e., $H_w = R^{-\frac{1}{2}}$) indeed results in a unity variance diagonal output noise covariance, that is, $\text{cov}(H_w \mathbf{n}) = I$, where $\text{cov}(\cdot)$ denotes the covariance operator. (Hint: H_w is not a random variable and thus is unaffected by the expectation operator.)

3. Assume a target has a transfer matrix given by

$$H_T = \begin{bmatrix} 1 & 1/\sqrt{2} \\ 1/\sqrt{2} & 1 \end{bmatrix}$$

- a. What is the optimum input (eigenfunction) that maximizes SNR for the white noise case?
 - b. If we interpret H_w as a target response polarization matrix in a H-V (horizontal and vertical) basis, what is the optimum polarization direction, assuming H-polarization corresponds to 0 degrees and V corresponds to 90 degrees?
4. The original formulation of equation (3.12) was in the analog domain [33, 34].
- a. Assuming the composite channel transfer function (impulse response) is given by the real valued LTI response $h(t)$, show that the optimum transmit function $s(t)$ that maximizes output SNR satisfies

$$\int_0^T s(\tau_2) K(\tau_2, \tau_1) d\tau_2 = \lambda s(\tau_1)$$

where

$$K(\tau_1, \tau_2) \triangleq \int_0^T h(t - \tau_1) h(t - \tau_2) dt$$

and where it is assumed that the pulse duration and receiver integration times are equal to T .

- b. Repeat assuming that the impulse response is complex valued.
5. Using MATLAB or some other suitable numerical processing software, compute the H matrices for the short- and long-pulse cases of example 3.1 and verify the results for the optimum transmit waveforms (short and long).
6. Repeat example 3.3 assuming that the target now spans two range bins, that is, $\delta[n] + \delta[n - 1]$. Does the result make sense in terms of minimizing interference from clutter?
7. For the two-target optimum ID problem, show that:
- a. Maximizing the norm of the separation metric $\mathbf{d} \triangleq \mathbf{y}_1 - \mathbf{y}_2$ in equation (3.42) is statistically optimum for the additive white noise case assuming a unimodal PDF and monotonic distribution function.
 - b. Extend this to the additive colored noise case (same PDF and distribution assumptions) and show that the separation metric to maximize is now the difference between the whitened target echo responses.
8. The energy in the whitened target echo for the infinite duration case is given by

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 |S(\omega)|^2 d\omega$$

where $Y(\omega)$, $H(\omega)$, $S(\omega)$ are the Fourier transforms of the whitened target echo, composite channel transfer function, and input (transmit) waveform, respectively. Show that the input $S(\omega)$ that maximizes the output energy satisfies $|S(\omega)| \propto |H(\omega)|$ [35].

9. A constrained optimum MIMO approach can be developed based on recognizing that the N -dimensional eigenspectrum of the generally positive definite composite channel kernel $H'H$ (or $E\{H'H\}$ for the stochastic case) forms a continuum for which some number k of eigenfunctions (and corresponding eigenvalues) retain matching properties.
- a. Assume that k orthonormal eigenfunctions of $H'H$, denoted by $\mathbf{u}_1, \dots, \mathbf{u}_k$, with associated eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$, are available and have better matching properties than, say, a nominal nonadaptive quiescent waveform \mathbf{s}_q . Derive an expression for the waveform \mathbf{s}_p that resides in the matched subspace spanned by the k best eigenvectors. The resulting waveform can be viewed as a type of constrained optimization in which the properties of the nominal waveform (e.g., good range sidelobes) are traded for better SNR (see, e.g., [36]).
 - b. Show that in the limit as $k \rightarrow N$, the matched subspace waveform $\mathbf{s}_p \rightarrow \mathbf{s}_q$. (Hint: The eigenfunctions of a positive definite (Hermitian) matrix form a complete basis.)

MIMO Radar

Michael S. Davis

Chapter Outline

4.1	Introduction	119
4.2	An Overview of MIMO Radar	121
4.3	The MIMO Virtual Array	122
4.4	MIMO Radar Signal Processing	124
4.5	Waveforms for MIMO Radar	135
4.6	Applications of MIMO Radar	138
4.7	Summary	142
4.8	Further Reading	143
4.9	References	143
4.10	Problems	145

4.1 INTRODUCTION

Radar systems have long used multiple antennas on receive to increase performance, but the idea of using multiple transmit antennas to emit a collection of independent waveforms is relatively new. Such a radar is called a multiple-input, multiple-output (MIMO) radar because it probes its environment by applying multiple inputs and observing multiple outputs. Inspired by the success of MIMO techniques at improving wireless communications, researchers have been actively investigating the potential of MIMO radar since the turn of the century [1].

In a sense, a MIMO radar is a natural extension of the phased array antenna that has been used by radar systems for decades [2]. A phased array consists of a number of radiating elements. Each element transmits the same signal except that a phase shift (or time delay) is applied so that a beam is steered in a particular direction. In a phased array, the signals transmitted from each element are perfectly correlated. A MIMO radar is a generalization of a phased array in that the signals need not be correlated from element to element.

4.1.1 Organization

First, a brief overview of the MIMO radar concept is provided. The utility of transmitting orthogonal waveforms is motivated by introducing the idea of the MIMO virtual array. A signal processing framework is then established, which provides insight into the

performance of a MIMO radar that transmits signals with arbitrary correlation properties. A brief discussion of MIMO radar waveforms is provided. Finally, applications of MIMO techniques to particular radar modes are described.

Some key points that will be discussed in this chapter include the following:

- A MIMO radar uses multiple transmit elements that emit independent waveforms and observes the returns from a scene of interest using multiple receive elements.
- A MIMO radar with closely spaced antennas is an extension of the traditional phased array. Standard array configurations preserve receive degrees of freedom by digitizing signals observed by multiple, spatially diverse channels. If a radar could transmit orthogonal waveforms with spatial diversity, transmit degrees of freedom would also be preserved, leading to a more flexible radar system.
- A MIMO radar with widely separated antennas is an extension of the bistatic radar concept. Each bistatic pair of radars observes a statistically independent realization of target reflectivity that may allow tracking to continue where it would otherwise be interrupted by target fading or an unfavorable geometry.
- The enhanced angular resolution provided by a MIMO radar can be understood by considering how, using multiple transmitted waveforms, a larger virtual array can be synthesized compared with using a single transmit phase center.
- The characteristics of a MIMO radar are described by the correlation between the transmitted signals. A phased array transmits perfectly correlated waveforms to form a narrow, high gain beam. By using uncorrelated waveforms, a MIMO radar trades off this peak gain for a more flexible antenna that can digitally re-steer its transmit beam and provide enhanced resolution.

4.1.2 Notation

The conjugate of a scalar, z , is denoted z^* . The Hermitian (conjugate) transpose of a matrix, \mathbf{A} , is denoted \mathbf{A}^H . The pseudoinverse of a matrix, \mathbf{A} , is denoted \mathbf{A}^+ . The Kronecker product of two matrices, \mathbf{A} and \mathbf{B} , is denoted $\mathbf{A} \otimes \mathbf{B}$. The $N \times N$ identity matrix is denoted \mathbf{I}_N .

Frequently used variables that are found in this chapter include

M = number of transmit elements/subarrays

N = number of receive elements/subarrays

θ = angle relative to the array ($\theta = 0$ is broadside)

θ_0 = angle used for digital beamforming

$\tilde{\theta}_0$ = angle used for analog beamforming

$\mathbf{a}(\theta)$ = transmit steering vector corresponding to angle θ (length: M)

$\mathbf{b}(\theta)$ = receive steering vector corresponding to angle θ (length: N)

$\mathbf{A}(\theta)$ = MIMO channel matrix (size: $N \times M$)

$\phi_m(t)$ = waveform used by transmitter m

$\Phi(t)$ = vector of waveforms used by each transmitter (length: M)

$\mathbf{y}(t; \theta)$ = vector of signals observed by each receiver for a target at angle θ (length: N)

$\eta(t)$ = vector of receiver noise signals (length: N)

\mathbf{R}_ϕ = MIMO signal correlation matrix (size: $M \times M$)

$\mathbf{z}(\theta)$ = vector of data samples in the target range bin for each transmit/receive pair after matched filtering for a target at angle θ (length: MN)

\mathbf{e} = vector of filtered receiver noise in the target range bin (length: MN)

$\mathbf{s}(\theta)$ = MIMO steering vector for angle θ (length: MN)

\mathbf{R}_e = MIMO noise covariance matrix (size: $MN \times MN$)

4.1.3 Acronyms

Commonly used acronyms that are found in this chapter include

ACR	area coverage rate
CDM	code division multiplexing
FDM	frequency division multiplexing
GMTI	ground moving target indication
LFM	linear frequency modulation
MIMO	multiple-input, multiple-output
MTI	moving target indication
PRF	pulse repetition frequency
PSF	point spread function
SAR	synthetic aperture radar
SIMO	single-input, multiple-output
SINR	signal-to-interference-plus-noise ratio
SISO	single-input, single-output
SNR	signal-to-noise ratio
TDM	time division multiplexing

4.2 | AN OVERVIEW OF MIMO RADAR

A MIMO radar is characterized by its ability to emit independent signals from multiple spatially diverse transmit elements and to observe the returns from multiple spatially diverse receive elements. The operation of a MIMO radar is illustrated in Figure 4-1. The idea of exploiting transmit diversity in a radar system is not new (e.g., see [3]), and similar concepts have been discussed in other contexts [4,5]; however, the theory developed in the past decade has leveraged the mathematical framework developed by the wireless communications community to characterize the performance of MIMO radar systems.

MIMO radars are typically divided into two classes depending on the separation between their elements. The distinction lies in whether we can assume that the returns from

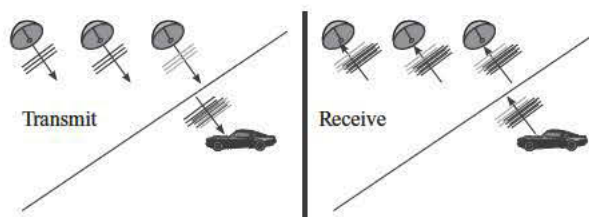


FIGURE 4-1 ■ MIMO radar signal model. Each radiating element transmits an independent signal. On receive, each element observes the echo of each transmitted signal.

each target will be correlated from element to element. Suppose that only a single target is illuminated and observed by the elements of a MIMO radar. If the elements are widely separated, perhaps distributed among a number of radar platforms, then each transmit/receive pair may observe uncorrelated echoes from the target due to aspect-dependent backscattering phenomena. On the other hand, if the elements are colocated on a single platform, they may be suitably close to observe the target from essentially the same aspect.

In a MIMO radar with widely separated antennas, each transmit/receive pair is assumed to observe an independent realization of the target reflectivity. This may be considered as an extension of the concept of a bistatic radar system, which consists of a single transmitter and a single receiver that are separated by a great distance, to include the possibility of using multiple transmitters and multiple receivers. This configuration has been called statistical MIMO, since it seeks to exploit this variation to improve detection and estimation performance. The use of widely spaced, multistatic systems was presented in the MIMO context in [6], but such techniques have a long history [7].

Assessing the utility of MIMO with widely separated antennas is not straightforward, since it relies on a system-level analysis to consider trade-offs between coverage rate and tracking utility since multiple radars must cooperate rather than operate independently. Also, many of the interesting challenges in realizing such systems are concerned with being able to share large volumes of data between distributed radar systems. For these reasons, MIMO radar with widely separated antennas will not be covered in the following discussion. A review of these techniques is provided in [8].

Instead, we will focus on MIMO radars in which the transmit/receive elements are closely spaced. We can therefore assume that the returns due to a particular target are correlated from element to element, which enables coherent processing across the MIMO array. For this reason, MIMO radars with widely separated antennas may be referred to as non-coherent MIMO while those with closely spaced antennas are called coherent MIMO radars.

Conventional array antenna technologies enable digital beamforming on receive by digitizing multiple spatial channels. This allows the receive beampattern to be resteeered, but, since the beampattern is the product of the transmit beampattern and the receive beampattern, this has limited utility; a beam cannot be formed in a direction in which no energy was transmitted. Using MIMO techniques, the transmit beampattern may be resteeered as well. Just as digitizing multiple channels on receive preserves receive degrees of freedom, transmitting orthogonal waveforms preserves transmit degrees of freedom. This concept of transmit beampattern flexibility will be developed in the following, as it is key to leveraging MIMO to enhance the capability of a radar system.

4.3 | THE MIMO VIRTUAL ARRAY

One way to interpret the ability of a MIMO radar to improve performance is to consider the virtual array corresponding to the physical array used by the radar. We will see how, by transmitting M orthogonal waveforms, a MIMO radar can synthesize an array that is M times larger than if a single transmit phase center is used. Just as using a physically larger antenna may provide a more desirable beampattern and improved resolution properties, synthesizing a larger virtual antenna provides these benefits as well.

An array antenna can be considered as a set of pseudo-bistatic systems (the term *bistatic* is typically reserved for radars that operate with transmitters and receivers that are

separated by great distances). In a sense, the information observed by an array antenna is the difference in path lengths between transmit/receive pairs as a function of angle of arrival. It can be shown, under reasonable assumptions, that a pseudo-bistatic transmit/receive pair is equivalent to a monostatic system that is located directly between the bistatic transmitter and receiver.

This is valid as long as the target is in the far field; that is, the range to the target is much greater than the size of the transmit and receive arrays. Let \mathbf{x}_T and \mathbf{x}_R denote the position of a transmit element and receive element, respectively, given in Cartesian coordinates relative to the center of the array. If \mathbf{u} is a unit vector that points from the array center to a target in the far field, and if the signal is narrowband relative to the size of the array, then the receiver at \mathbf{x}_R observes essentially the same signal as would be observed at the array center up to a phase shift. For this transmit/receive pair, the observed phase shift is

$$\frac{4\pi}{\lambda} \mathbf{u}^T \left(\frac{\mathbf{x}_T + \mathbf{x}_R}{2} \right)$$

where λ is the wavelength corresponding to the radar center frequency. Observe that this phase shift is the same as would be observed if the transmitter and receiver were both located at $(\mathbf{x}_T + \mathbf{x}_R)/2$, the point directly between them.

The set of all virtual elements that correspond to a set of physical transmit/receive elements is called the virtual array. This is similar to the coarray described in [4]. If the M transmit elements of a MIMO radar have positions described by $\mathbf{x}_{T,1}, \dots, \mathbf{x}_{T,M}$ and the N receive elements by $\mathbf{x}_{R,1}, \dots, \mathbf{x}_{R,N}$, then the virtual array is the set of positions

$$\left\{ \frac{\mathbf{x}_{T,m} + \mathbf{x}_{R,n}}{2} : m = 1, \dots, M; n = 1, \dots, N \right\} \quad (4.1)$$

In general, if M transmitters (using M orthogonal waveforms) are used and N receive channels are used, then the number of virtual phase centers will be MN . These virtual phase centers may not all be distinct depending on the array topology.

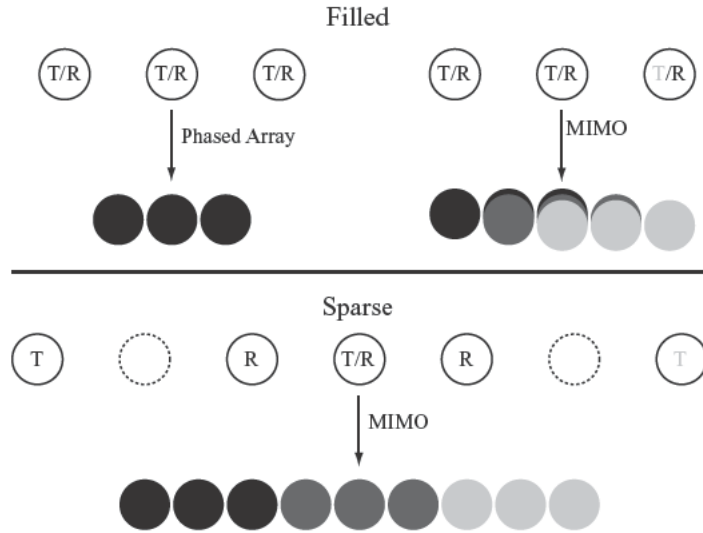
In a phased array, there is essentially one transmit element since the transmitted signals are perfectly correlated. After transmission, the radar cannot differentiate between signals transmitted from different elements. So, the virtual array is generated only by a single physical transmit element, as in the upper left of Figure 4-2. Note that the length of the virtual array is about half of the length of the physical phased array.

Now, suppose that the radar has multiple transmit elements and that each is capable of transmitting a waveform that is orthogonal to all of the others. Two waveforms, $\phi_1(t)$ and $\phi_2(t)$, are said to be orthogonal if

$$\int_{-\infty}^{\infty} \phi_1(t) \phi_2^*(t) dt = 0 \quad (4.2)$$

Essentially, a filter matched to the waveform ϕ_1 will completely reject any contribution due to ϕ_2 and vice versa. Since the transmitted signals are assumed to be orthogonal, the receiver is able to process the contribution of each transmitter independently; by applying a filter matched to a particular transmitted waveform, the contributions from all of the other transmitted waveforms are removed due to their orthogonality. The impact of correlations between the waveforms when they are not perfectly orthogonal will be discussed in subsequent sections.

FIGURE 4-2 ■
Examples of physical arrays and corresponding virtual arrays. In all of these cases, there are $M = 3$ transmit elements and $N = 3$ receive elements.



In this way, the same array can synthesize the larger virtual array in the upper right of Figure 4-2 by transmitting orthogonal instead of correlated waveforms. Note that some of the virtual elements overlap. The sparse array at the bottom of the figure shows an example where the virtual array is both filled and has no redundancy. This configuration is called a Nyquist array in [9], since it is critically sampled in the spatial sense.

4.4 | MIMO RADAR SIGNAL PROCESSING

While a distinction is made in the literature between single-input, single-output (SISO) and single-input, multiple-output (SIMO) systems and the general case of a MIMO system, all may be analyzed using a unified framework. This allows the phased array (SIMO) and orthogonal waveform cases to be treated as special cases of a MIMO radar. A signal model will be developed to accommodate these different cases where each will be characterized by its transmitted signal correlation matrix. A procedure for digital beamforming will be derived, and the resulting characteristics will be presented.

4.4.1 Signal Model

Consider a radar system that uses an array antenna for transmit and receive. Suppose that this array is partitioned into a number of subarrays. A subarray could be a portion of an active electronically scanned array (AESA) or simply a directional horn or a reflector antenna. Let M be the number of transmit subarrays and N be the number of receive subarrays.

For a particular pulse, the signal transmitted by subarray m is denoted $\phi_m(t)$. We will assume that the total transmitted energy is unity for a given pulse, that is,

$$\sum_{m=1}^M \int_{-\infty}^{\infty} |\phi_m(t)|^2 dt = 1 \quad (4.3)$$

The peak transmit power of the radar and other elements of the radar range equation will be accounted for elsewhere. Each of these M transmitted waveforms will reflect off targets in the scene, and the echoes will be observed by each of the N receivers.

If the narrowband assumption is valid (i.e., the array is small relative to the bandwidth of the signal), then each subarray observes the same signal up to a phase shift. It is the information encoded in this phase shift from subarray to subarray that allows beamforming to be implemented. For a particular wavelength and for a particular angle of interest θ , we can form the corresponding steering vector that describes these phase shifts. In the MIMO case, we will need two steering vectors: the length- M transmit steering vector, $\mathbf{a}(\theta)$; and the length- N receive steering vector, $\mathbf{b}(\theta)$. Assume that each element of the steering vectors is unit-magnitude. Consequently, we have

$$\|\mathbf{a}(\theta)\|^2 = M, \|\mathbf{b}(\theta)\|^2 = N \quad (4.4)$$

Note that, for this analysis, we assume that the signals are narrowband, but this model could be extended to include wideband effects.

We now seek a model for the data observed from a single point target. Consider a target at an angle θ that has a signal-to-noise ratio (SNR) of γ and a bulk phase shift of ψ . The signal observed by receiver n due to this target may be written

$$y_n(t; \theta) = \sqrt{\gamma} e^{i\psi} b_n(\theta) \sum_{m=1}^M a_m(\theta) \phi_m(t) + \eta_n(t) \quad (4.5)$$

where $a_m(\theta)$ and $b_n(\theta)$ are, respectively, elements of the transmit steering vector $\mathbf{a}(\theta)$ and the receive steering vector $\mathbf{b}(\theta)$ corresponding to the angle θ , and $\eta_n(t)$ is receiver noise. Of course, a radar operates in three-dimensional space, which requires that two angles be specified. In this treatment, only one angle, θ , is considered for clarity.

This data model includes only a single target and assumes that the processor has knowledge of the range of the target. In general, this is not a practical assumption, since one purpose of a radar system is to measure the range to a target. This is accomplished by hypothesizing a number of target ranges and processing each range bin independently. So, if a target is present in a particular range bin, then the assumption of known range is meaningful.

The signal presented in (4.5) describes the observation of one of the N receive subarrays. The data observed by all of the receivers may be written more compactly by grouping the data so that each element of a (column) vector corresponds to a receive subarray. Let $\Phi(t)$ be a (column) vector where each of the M elements is the waveform transmitted by a particular subarray.

$$\Phi(t) \triangleq \begin{pmatrix} \phi_1(t) \\ \vdots \\ \phi_M(t) \end{pmatrix} \quad (4.6)$$

The signal observed by a MIMO radar corresponding to a target at an angle θ is written

$$\mathbf{y}(t; \theta) = \sqrt{\gamma} e^{i\psi} \mathbf{b}(\theta) \mathbf{a}(\theta)^T \Phi(t) + \boldsymbol{\eta}(t) \quad (4.7)$$

where \mathbf{b} and \mathbf{a} are the receive and transmit steering vectors, respectively. We can define the MIMO channel matrix, $\mathbf{A}(\theta)$, in terms of the transmit steering vector, $\mathbf{a}(\theta)$, and the

where the integral of the matrix is computed element-wise. Each element of this matrix describes the correlation between two of the transmitted waveforms. As will be discussed later, this matrix characterizes the performance attributes of a MIMO radar.

The $N \times M$ data matrix given in (4.10) may be vectorized by stacking the columns of \mathbf{Z} . This yields

$$\begin{aligned} \mathbf{z}(\theta) &\triangleq \text{Vec}\{\mathbf{Z}(\theta)\} \\ &= \sqrt{\gamma} e^{i\psi} \left(\mathbf{R}_\phi^T \otimes \mathbf{I}_N \right) \text{Vec}\{\mathbf{A}(\theta)\} + \mathbf{e} \end{aligned} \quad (4.12)$$

where $\mathbf{e} \triangleq \text{Vec}\{\mathbf{E}\}$, and the operator \otimes is the Kronecker product. The following, well-known relationship between the vectorization operator and the Kronecker product was used [10]: if \mathbf{A} is an arbitrary $N \times K$ matrix and \mathbf{B} is an arbitrary $K \times M$ matrix, then we have

$$\begin{aligned} \text{Vec}\{\mathbf{AB}\} &= (\mathbf{I}_M \otimes \mathbf{A}) \text{Vec}\{\mathbf{B}\} \\ &= (\mathbf{B}^T \otimes \mathbf{I}_N) \text{Vec}\{\mathbf{A}\} \end{aligned} \quad (4.13)$$

We can define the MIMO steering vector, $\mathbf{s}(\theta)$, for an angle θ in terms of the MIMO channel matrix, $\mathbf{A}(\theta)$

$$\mathbf{s}(\theta) \triangleq \left(\mathbf{R}_\phi^T \otimes \mathbf{I}_N \right) \text{Vec}\{\mathbf{A}(\theta)\} \quad (4.14)$$

The covariance matrix of the noise vector is

$$\begin{aligned} \mathbf{R}_e &\triangleq \text{E}[\mathbf{e}\mathbf{e}^H] \\ &= \mathbf{R}_\phi^T \otimes \mathbf{I}_N \end{aligned} \quad (4.15)$$

These definitions allow us to write the MIMO signal model after matched filter processing as

$$\mathbf{z}(\theta) = \sqrt{\gamma} e^{i\psi} \mathbf{s}(\theta) + \mathbf{e} \quad (4.16)$$

where the MIMO steering vector is given in (4.14) and the noise covariance matrix in (4.15).

If the radar transmits orthogonal waveforms each of unit-power, then the MIMO signal correlation matrix is an identity matrix ($\mathbf{R}_\phi = \mathbf{I}$). For this special case, the data vector is

$$\mathbf{z}_\perp(\theta) = \sqrt{\gamma} e^{i\psi} \text{Vec}\{\mathbf{A}(\theta)\} + \mathbf{e} \quad (4.17)$$

where the covariance matrix of the noise is spatially white. By transmitting orthogonal waveforms, the radar retains its transmit degrees of freedom in that it has a full observation of the full MIMO steering matrix, $\mathbf{A}(\theta)$.

It is obvious from this signal model and the MIMO steering vector of (4.14) that the MIMO signal correlation matrix, \mathbf{R}_ϕ , characterizes the performance of a MIMO radar.

4.4.2 MIMO Signal Correlation Matrix

As is evident from the definition in (4.11) and the preceding discussion, the MIMO signal correlation matrix, \mathbf{R}_ϕ , is the $M \times M$ matrix that describes the (zero-lag) cross- and

auto-correlations of the M transmitted waveforms, $\{\phi_m(t)\}$. The element in row m and column m' describes the response of waveform m when the matched filter corresponding to waveform m' is applied. This element of the matrix may be written

$$(\mathbf{R}_\phi)_{m,m'} \triangleq \int_{-\infty}^{\infty} \phi_m(t) \phi_{m'}^*(t) dt \quad (4.18)$$

Since a phased array is a special case of MIMO radar, it also possesses a MIMO signal correlation matrix. Let $\phi_0(t)$ be the radar waveform used by a phased array system. Each subarray will transmit this signal but with a phase shift applied to steer the beam in a particular direction. To steer a beam in the direction $\tilde{\theta}_0$, the transmitted signals may be written

$$\Phi_{PA}(t) \triangleq \mathbf{a}^*(\tilde{\theta}_0) \phi_0(t) \quad (4.19)$$

where $\mathbf{a}(\theta)$ is the transmit steering vector corresponding to the direction θ . To satisfy the energy constraint of (4.3), the signal $\phi_0(t)$ must be normalized so that

$$\int_{-\infty}^{\infty} |\phi_0(t)|^2 dt = \frac{1}{M} \quad (4.20)$$

The signal correlation matrix is found to be

$$\mathbf{R}_{\phi/PA} = \frac{1}{M} \mathbf{a}^*(\tilde{\theta}_0) \mathbf{a}^*(\tilde{\theta}_0)^H \quad (4.21)$$

This steering angle, $\tilde{\theta}_0$, must be chosen by the phased array before transmitting since this is a form of analog beamforming. As the phased array scans its transmit beam from dwell to dwell, the signal correlation matrix will change, but it will remain rank-1.

Note that the correlation matrix is scaled so that the trace (the sum of the diagonal elements) of the correlation matrix is 1. This is required so that the signal-to-noise ratio remains γ regardless of the structure of the MIMO correlation matrix. In effect, this enforces a constant transmit power between designs to allow reasonable comparisons.

Consider now the case where each subarray transmits one of a suite of orthogonal waveforms; the correlation matrix is full-rank. If orthogonal waveforms are used, each with equal power, then the correlation matrix is a scaled identity matrix,

$$\mathbf{R}_{\phi/\perp} = \frac{1}{M} \mathbf{I}_M \quad (4.22)$$

Another example is when the array is spoiled on transmit. To cover a larger area, the transmit beamwidth may be increased by applying a phase taper across the array or by transmitting out of a single subarray. In the latter case, the transmitted waveforms are considered to be identically zero for all but one subarray. This approach is referred to as spoiling on transmit since it effectively spoils the transmit beampattern of the phased array by trading peak gain for a wider coverage area. The signal correlation matrix of the spoiled phased array, $\mathbf{R}_{\phi/Spoil}$, is rank-1, just as in the unspoiled phased array case. In the context of synthetic aperture radar (SAR), as well as in synthetic aperture sonar, an array that uses a single element on transmit and multiple elements on receive is referred to as a Vernier array.

Consider the case of three transmitters ($M = 3$). The MIMO signal correlation matrices for the phased array (steered to broadside), the spoiled phased array, and the radar using orthogonal waveforms are

$$\mathbf{R}_{\phi/PA} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \mathbf{R}_{\phi/Spoil} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{R}_{\phi/\perp} = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}$$

By considering the extreme cases of the phased array (rank-1 correlation matrix) and the radar using orthogonal waveforms (full-rank correlation matrix), the effect of the signal correlation matrix on the energy transmitted can be observed. As is well-known, a phased array radar transmits most of its energy in a particular direction. Of course, to accomplish this without violating laws of physics, energy is not transmitted in other directions. On the other hand, a radar transmitting orthogonal waveforms emits energy in all directions with similar power (subject to the pattern of the subarray).

An example of the radiated power as a function of angle comparing the phased array and the radar using orthogonal waveforms is presented in Figure 4-4. As expected, the phased array transmits a set of waveforms that cohere in a desired direction and destructively interfere in other directions. The spoiled phased array uses a single (omnidirectional) element, so no beamforming gain is provided. Similarly, the radar using orthogonal waveforms transmits uncorrelated waveforms. Unlike the phased array, these signals do not cohere in any preferred direction.

There is a continuum of signal correlation matrices between these extremes of completely uncorrelated and perfectly correlated waveforms. It will be shown that, by choosing the correlation among the signals transmitted by a MIMO radar, peak gain can be traded for a larger access area through transmit restearing.

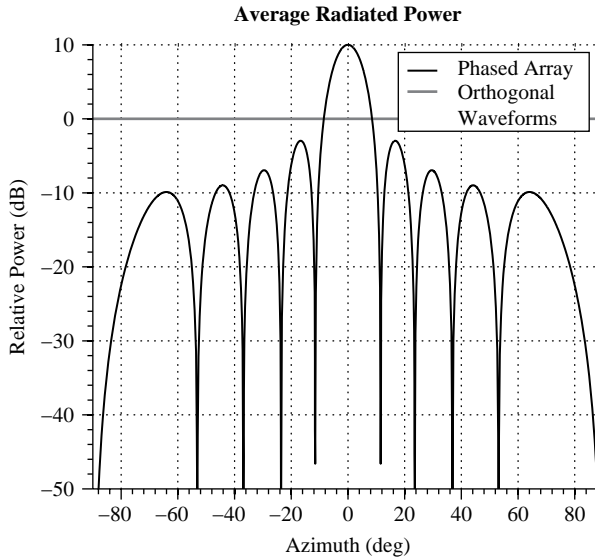


FIGURE 4-4 ■ Average radiated power as a function of angle. Assume that $M = 10$ elements are used on transmit, the elements are spaced at half-wavelength intervals, and each element is omnidirectional.

4.4.3 MIMO Spatial Beamforming

A model has been developed, given by (4.16), that describes the characteristics of the signal observed by a MIMO radar due to a target at a particular angle, θ . This can be used to construct a spatial beamformer that selects signals due to targets at a direction of interest, θ_0 , while attempting to reject signals due to targets arriving from other angles.

Let $\mathbf{w}(\theta_0)$ be the spatial weights that are chosen to steer a beam in the direction θ_0 . The output of the linear beamformer described by the weights $\mathbf{w}(\theta_0)$ when a target is present at an angle θ may be written

$$\mathbf{w}(\theta_0)^H \mathbf{z}(\theta) = \underbrace{\sqrt{\gamma} e^{i\psi} \mathbf{w}(\theta_0)^H \mathbf{s}(\theta)}_{\text{Signal}} + \underbrace{\mathbf{w}(\theta_0)^H \mathbf{e}}_{\text{Noise}} \quad (4.23)$$

The goal is to choose a set of spatial weights, \mathbf{w} , that maximizes the output signal-to-noise ratio.

$$\text{SNR}_{out} = \gamma \frac{\left| \mathbf{w}(\theta_0)^H (\mathbf{R}_\phi^T \otimes \mathbf{I}_N) \text{Vec} \{ \mathbf{A}(\theta) \} \right|^2}{\mathbf{w}(\theta_0)^H (\mathbf{R}_\phi^T \otimes \mathbf{I}_N) \mathbf{w}(\theta_0)} \quad (4.24)$$

It is well-known (see [11]) that the optimal weights, $\mathbf{w}(\theta_0)$, satisfy the equations

$$\mathbf{R}_e \mathbf{w}(\theta_0) = \mathbf{s}(\theta_0) \quad (4.25)$$

where the MIMO steering vector, \mathbf{s} , and the noise covariance matrix, \mathbf{R}_e , were given in (4.14) and (4.15), respectively.

The solution of this optimization problem is straightforward so long as \mathbf{R}_e is full-rank, but this is the case only when the MIMO signal correlation matrix, \mathbf{R}_ϕ , is full-rank. As was seen in the previous section, in the case of the phased array, the MIMO signal correlation matrix is rank-1, so (except for the degenerate case of $M = 1$) no unique set of spatial weights is optimal since there is an infinite set of spatial weights that achieve the same output SNR.

When the signal correlation matrix is not full-rank, the matrix \mathbf{R}_e has a nontrivial null space. If, for a particular angle of interest, the vectorized steering matrix $\text{Vec} \{ \mathbf{A}(\theta_0) \}$, lies in this null space, then the MIMO steering vector is zero and the constrained optimization problem cannot be solved; a beam cannot be formed in the direction θ_0 .

It is not surprising that a beam can be formed in every direction only when the transmitted signals are orthogonal. This is because when the signals are correlated with each other, like in a phased array, energy is not radiated in some directions corresponding to nulls of the transmit pattern. It is precisely at these angles where no energy was transmitted that no beam can be formed.

If a beam can be formed in the direction θ_0 , the appropriate spatial weights are

$$\mathbf{w}(\theta_0) = \text{Vec} \left\{ \mathbf{A}(\theta_0) \mathbf{R}_\phi^+ \mathbf{R}_\phi^+ \right\} \quad (4.26)$$

where \mathbf{X}^+ denotes the pseudoinverse of a matrix, \mathbf{X} [12]. If a matrix is full-rank, then its inverse and pseudoinverse coincide, in which case the appropriate spatial weights are simply the vectorized MIMO channel matrix, $\mathbf{A}(\theta_0)$, corresponding to the angle of interest, θ_0 .

4.4.4 MIMO Gain

The array factor describes the pattern of an array antenna if each subarray was omnidirectional. The array factor of a MIMO radar in the direction θ when the beamforming weights are steered to the direction θ_0 is defined by

$$f(\theta; \theta_0) \triangleq \frac{\mathbf{w}(\theta_0)^H \mathbf{s}(\theta)}{\sqrt{\mathbf{w}(\theta_0)^H \mathbf{R}_e \mathbf{w}(\theta_0)}} \quad (4.27)$$

When the optimal weights are used, this may be written in terms of the MIMO steering matrix \mathbf{A} or, alternatively, in terms of the transmit/receive steering vectors \mathbf{a} and \mathbf{b} .

$$f(\theta; \theta_0) = \frac{\text{Vec}\{\mathbf{A}(\theta_0)\}^H (\mathbf{R}_\phi^T \otimes \mathbf{I}_n) \text{Vec}\{\mathbf{A}(\theta)\}}{\sqrt{\text{Vec}\{\mathbf{A}(\theta_0)\}^H (\mathbf{R}_\phi^T \otimes \mathbf{I}_n) \text{Vec}\{\mathbf{A}(\theta_0)\}}} \quad (4.28)$$

$$= \left(\frac{\mathbf{a}(\theta_0)^H \mathbf{R}_\phi^T \mathbf{a}(\theta)}{\sqrt{\mathbf{a}(\theta_0)^H \mathbf{R}_\phi^T \mathbf{a}(\theta_0)}} \right) \left(\frac{\mathbf{b}(\theta_0)^H \mathbf{b}(\theta)}{\sqrt{N}} \right) \quad (4.29)$$

In general, the array factor will be complex valued. The gain of an antenna is the magnitude-squared of the array pattern in addition to the gain of the subarrays. The antenna gain $G(\theta; \theta_0)$ describes the increase in SNR that a target at angle θ will receive if a beam is digitally steered in the direction θ_0 .

If it is assumed that all of the transmit subarrays and receive subarrays of a MIMO array antenna are identical, then the (two-way) gain of the MIMO radar is

$$G(\theta; \theta_0) = \underbrace{\left(E_{TX}(\theta) \frac{|\mathbf{a}(\theta_0)^H \mathbf{R}_\phi^T \mathbf{a}(\theta)|^2}{\mathbf{a}(\theta_0)^H \mathbf{R}_\phi^T \mathbf{a}(\theta_0)} \right)}_{\text{TransmitGain}} \underbrace{\left(E_{RX}(\theta) \frac{|\mathbf{b}(\theta_0)^H \mathbf{b}(\theta)|^2}{N} \right)}_{\text{ReceiveGain}} \quad (4.30)$$

where $E_{TX}(\theta)$ and $E_{RX}(\theta)$ are the (one-way) subarray gains on transmit and receive, respectively. It is important to observe that the transmit gain is strongly dependent on the signal correlation matrix, \mathbf{R}_ϕ . A similar development is presented in [13].

4.4.5 Phased Array versus Orthogonal Waveforms

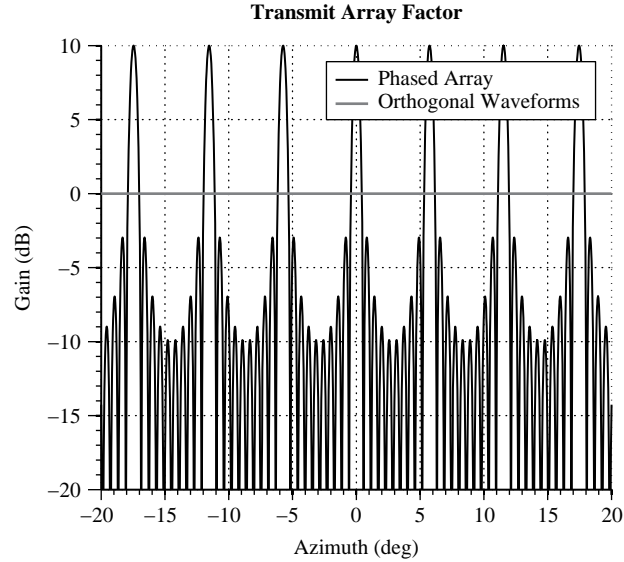
We have shown that the transmit gain of an array antenna can be controlled by design of the MIMO signal correlation matrix. The two extreme cases are considered: the phased array (rank-1 matrix) and orthogonal waveforms (full-rank matrix). These matrices were given in (4.21) and (4.22).

The gain for these two cases can be computed using (4.30). The gains of the phased array, G_{PA} , and of the radar using orthogonal waveforms, G_\perp , are found to be

$$G_{PA}(\theta; \tilde{\theta}_0, \theta_0) = \left(E_{TX}(\theta) \frac{|\mathbf{a}(\tilde{\theta}_0)^H \mathbf{a}(\theta)|^2}{M} \right) \left(E_{RX}(\theta) \frac{|\mathbf{b}(\theta_0)^H \mathbf{b}(\theta)|^2}{N} \right) \quad (4.31)$$

$$G_\perp(\theta; \theta_0) = \left(E_{TX}(\theta) \frac{|\mathbf{a}(\theta_0)^H \mathbf{a}(\theta)|^2}{M^2} \right) \left(E_{RX}(\theta) \frac{|\mathbf{b}(\theta_0)^H \mathbf{b}(\theta)|^2}{N} \right) \quad (4.32)$$

FIGURE 4-5 ■
MIMO array factor on transmit. This example corresponds to an array with 10 subarrays where each subarray has $20 \lambda/2$ spaced elements.



Recall that θ_0 is the angle to which the beam is digitally steered and that $\tilde{\theta}_0$ is the direction to which the phased array steered the beam on transmit. The receiver is free to vary θ_0 with digital processing, but $\tilde{\theta}_0$ is fixed.

As should be expected, the receive gains are identical between the phased array and orthogonal waveforms, but there are two key differences between the transmit gain terms. First, since the phased array transmits a concentrated beam in the direction $\tilde{\theta}_0$, it is unable to apply any digital steering of the transmit beam; the phased array has already decided in which direction to send energy. On the other hand, the radar that employs orthogonal signals is able to re-steer the transmit beam to any angle (so long as the subarray pattern permitted energy to be radiated in that direction).

However, the cost of doing this is evident. The phased array realizes a transmit beam-forming gain that provides an increase in SNR by a factor of M , the number of transmitting subarrays. This benefit is lost by the radar that uses orthogonal signals.

These differences are illustrated in Figure 4-5, where the array factors of a MIMO radar employing orthogonal waveforms as well as a traditional phased array are presented.

The performance of an array antenna for use in a radar system is well quantified by considering three gain patterns: the steered response, the beampattern, and the point spread function. These describe the ability of the data collected by the system to be used to digitally form beams in desired directions with desired properties.

The steered response and the beampattern quantify the degree to which the antenna can be digitally steered to an angle of interest as well as the ability to reject returns from undesired angles. Given an angle of interest, the steered response describes the ability of the array to observe signals arriving from that direction when the array is steered to that direction of interest, while the beampattern describes the ability of the array to reject targets from other angles [14]. The distinctions between these patterns are summarized in Table 4-1.

Let $G_{RX}(\theta; \theta_0)$ be the gain of the receive array in the direction θ when it is digitally steered to the angle θ_0 . The steered response evaluates this gain for the case when $\theta = \theta_0$, that is, the gain is evaluated in the direction that the array has been digitally steered. If the array is steered to the angle θ_0 , then the beampattern evaluated at θ describes how much

TABLE 4-1 ■ Description of antenna patterns

Quantity	Description	Definition
Steered response	Ability to digitally re-steer antenna	$G_1(\theta) = \{G(\theta; \theta) : \theta \in \Theta\}$
Beampattern	Ability to reject targets from undesired angles	$G_2(\theta; \theta_0) = \{G(\theta; \theta_0) : \theta \in \Theta\}$ for fixed θ_0
Point spread function	Ability to resolve closely spaced targets	$G_3(\theta_0; \theta) = \{G(\theta; \theta_0) : \theta_0 \in \Theta_0\}$ for fixed θ

energy is observed from this direction θ . Note that a different beampattern is provided for each steered angle θ_0 .

The steered response and beampattern of a radar employing a phased array are compared to those of a system using orthogonal waveforms in Figure 4-6. It can be seen from the steered responses that, by using orthogonal waveforms, the radar is able to access a much larger area by re-steering its transmit beam, but this comes at a cost of peak gain. This can also be achieved by using the established technique of spoiling the beam on transmit through using a single subarray on transmit. The advantage of orthogonal waveforms is evident from the beampattern. By transmitting orthogonal waveforms, a radar can realize the same steered response offered by spoiling on transmit, but it can preserve the beampattern of the unspoiled phased array.

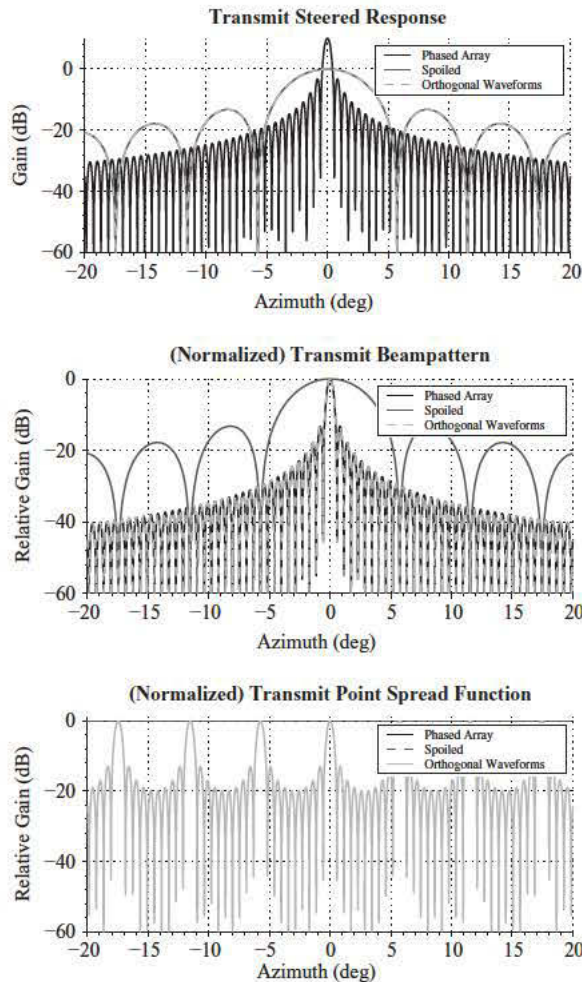
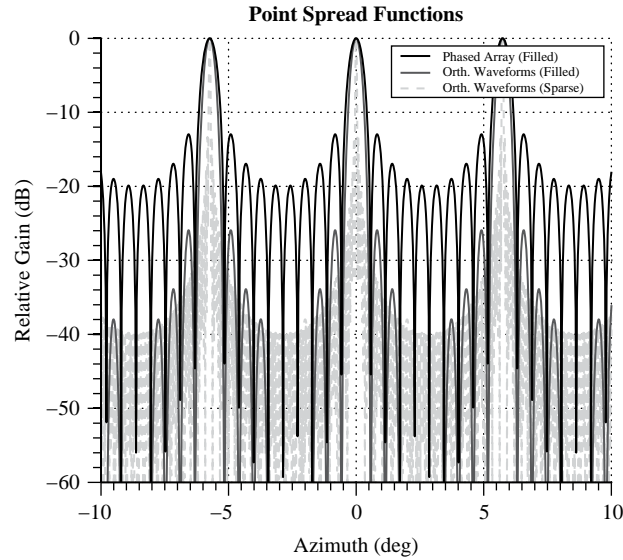


FIGURE 4-6 ■ Comparison of phased array and orthogonal waveforms. The subarray size is 10λ . Each configuration uses $M = 10$ transmit subarrays.

FIGURE 4-7 ■

Two-way angular point spread functions. The phased array resolves targets in angle using only receive degrees of freedom. For the orthogonal waveform cases, the filled configuration provides enhanced sidelobe performance and the sparse configuration provides improved angular resolution.



The (angular) point spread function (PSF) quantifies the angular resolution capability of an array antenna system. Suppose that a target is present and located at some angle θ . To evaluate the PSF at angle θ_0 , a beamformer designed for this angle of interest θ_0 is applied to data containing a target at angle θ . This quantifies the degree to which a target located at θ will obscure a target located at θ_0 .

Comparing the transmit PSF of the phased array to the orthogonal waveforms in Figure 4-6, we see that the phased array (regardless of spoiling) provides no angular resolution on transmit. If, on the other hand, orthogonal waveforms are used and the radar can preserve its transmit degrees of freedom, then angular resolution is possible on transmit.

The PSF is related to the ambiguity function that is familiar from the radar literature. The standard radar ambiguity function describes the response due to a target with a particular range and Doppler in nearby range and Doppler bins. This idea was extended to the MIMO case in [15], where an ambiguity function in terms of range, Doppler, and angle is developed.

Let us now reconsider the sparse configuration presented in Figure 4-2. The (two-way) PSF for the sparse array using orthogonal waveforms is presented in Figure 4-7. As predicted by the virtual array analysis, the sparse array is able to provide enhanced resolution since it provides an effectively larger array than the filled configuration. Note that orthogonal waveforms are required to use the sparse configuration; otherwise, the sparsity would introduce undesirable grating lobes. The filled configuration provides improved sidelobe performance, since a taper is effectively applied to the aperture as a result of overlapping virtual phase centers.

These results demonstrate the ability of a radar to use orthogonal waveforms to improve performance. Even if truly orthogonal waveforms are not practical, this analysis may be repeated for a given set of waveforms using the appropriate signal correlation matrix to quantify the impact of this nonorthogonality with the framework that has been developed.

4.5 | WAVEFORMS FOR MIMO RADAR

We have seen how the choice of transmitted waveforms determines the characteristics of a MIMO radar by considering the correlation matrix, \mathbf{R}_ϕ . The challenge of MIMO radar is to design families of waveforms that possess a desired correlation matrix. Also, note that the previous analysis considered only the zero-lag correlation matrix. By extending this analysis to consider the correlation matrix as a function of delay, we can study the behavior of the range sidelobes. An open research topic is to identify waveforms that enable the benefits of MIMO radar without unacceptably compromising other aspects of radar performance.

4.5.1 Classes of Waveforms for MIMO Radar

Three techniques exist to minimize the cross-correlation among a suite of waveforms, which correspond to exploiting orthogonality in time, frequency, and/or code.

Time Division Multiplexing (TDM) An obvious method to decorrelate waveforms is to simply transmit them at different times. This is possible in some applications, but in others it may not be compatible with requirements of the radar timeline. This will also increase data handling requirements since each transmitted waveform must be digitized separately.

Frequency Division Multiplexing (FDM) Another natural way to limit correlation between waveforms is to offset them in frequency. The drawback of this is apparent in systems that rely on coherent processing across a number of frequencies. In such cases, the target reflectivities may vary if the chosen frequency is too large, limiting coherent gain. Also, coherent imaging techniques like SAR rely on each spatial sample having the same frequency support; otherwise, artifacts will be present in the image.

Code Division Multiplexing (CDM) In wireless communications, it is frequently necessary for a number of users to access a particular frequency band at the same time. A common solution is to apply code division multiple access (CDMA) techniques. Even though each user transmits at the same time and at the same frequency, each signal is modulated by a unique phase code.

Another approach for CDM is to employ noise like waveforms. These have the drawback that they may have a high peak-to-average power-ratio, which requires transmit amplifiers that are more linear and thus have lower gain.

The synthesis of waveforms that are reasonably orthogonal that possess desired sidelobe responses is a key challenge in the realization of a MIMO radar system. It is important to realize that different radar applications have different sidelobe requirements. For example, in moving target indication (MTI) radar, the peak sidelobe ratio is paramount. This is contrasted with applications where a continuum of scatterers is expected, as in SAR and weather radar, where the integrated sidelobe ratio tends to drive performance.

Some may argue that TDM and FDM are not properly MIMO waveforms. They are included here to emphasize that MIMO radar is a generalization of traditional radar systems. For example, a number of radars operating cooperatively using difference frequencies may be considered as a single MIMO radar system. The utility of developing a theory of MIMO radar is to discuss this and many other configurations using a common framework. Including these waveforms greatly expands the class of MIMO radars.

4.5.2 MIMO Range Response

In the previous section, we described the angular point spread function of a MIMO radar, which described the sidelobes of the effective antenna pattern. Similar sidelobe structures will exist in the range domain, just as they do in traditional radar systems.

We extend the definition of the MIMO signal correlation matrix of (4.11) to

$$\mathbf{R}_\phi(\tau) \triangleq \int_{-\infty}^{\infty} \boldsymbol{\Phi}(t) \boldsymbol{\Phi}(t - \tau)^H dt \quad (4.33)$$

Note that our analysis of the spatial characteristics of a MIMO radar focused on the case where $\tau = 0$, the range bin of interest. This dependence on lag, τ , will allow us to characterize the influence a target at a particular range will have in bins up- and down-range.

The range response of a set of MIMO waveforms with correlation matrix $\mathbf{R}_\phi(\tau)$ is given by

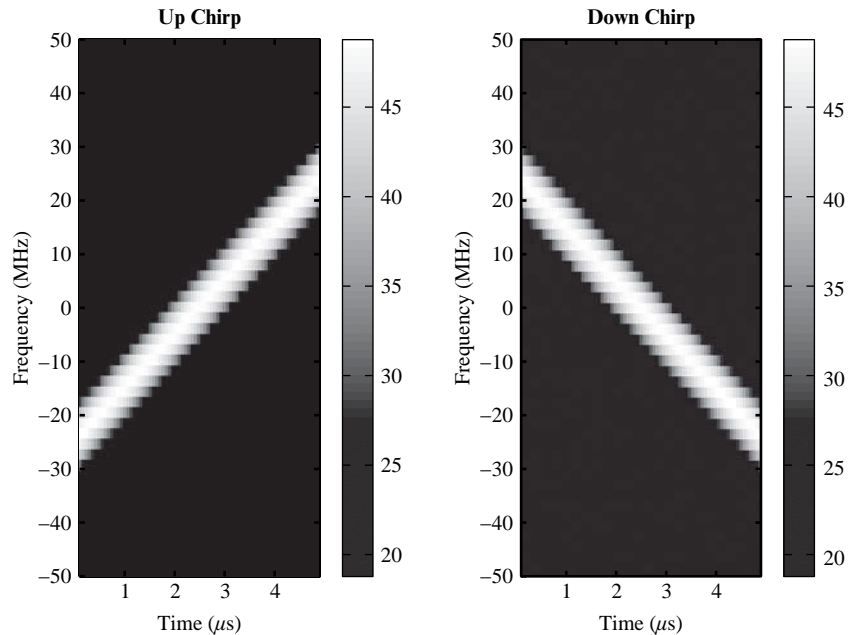
$$h(t, \theta; \theta_0) = \frac{\mathbf{a}(\theta_0)^H \mathbf{R}_\phi^*(t) \mathbf{a}(\theta)}{\sqrt{\mathbf{a}(\theta_0)^H \mathbf{R}_\phi^*(0) \mathbf{a}(\theta_0)}} \quad (4.34)$$

This result is extended to consider a Doppler offset in addition to a range offset in [15].

4.5.3 Example: Up- and Down-Chirp

Since many radar systems use linear frequency modulated (LFM) waveforms, also called an LFM chirp, a natural method for generating two approximately orthogonal waveforms is to use one chirp with increasing frequency and another with decreasing frequency. Time-frequency representations describing a notional pair of such signals is shown in Figure 4-8. Note that these signals possess the same frequency support.

FIGURE 4-8 ■ Time-frequency representation of the up- and down-chirps. In each case, the bandwidth is 50 MHz and the pulse width is 5 μ s.



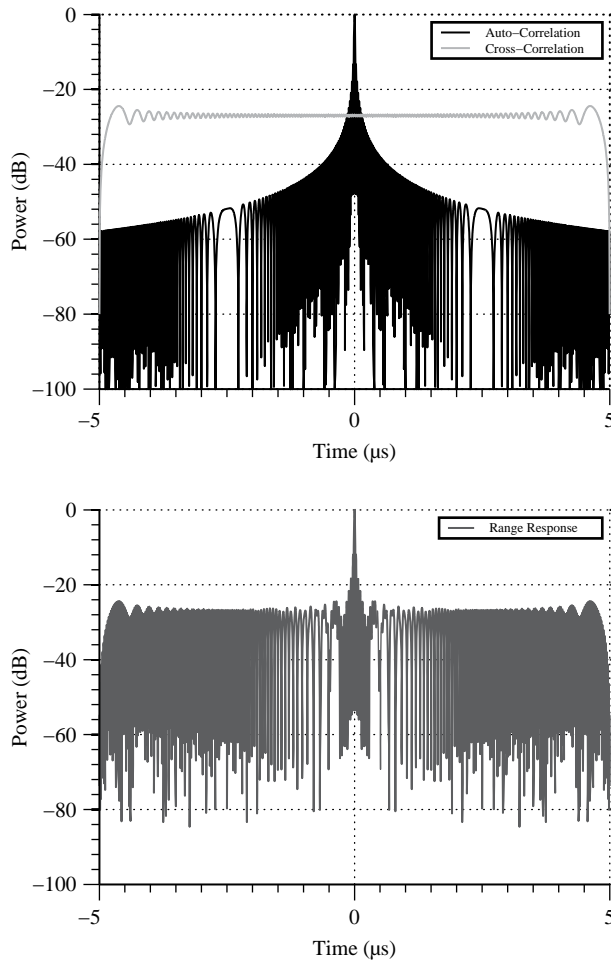


FIGURE 4-9 ■ The correlation properties and range response of the up- and down-chirp waveforms.

As shown in Figure 4-9, each waveform has a desirable auto-correlation, and the peak of the cross-correlation is well below that of the auto-correlation. However, the cross-correlation does not decay as the delay offset increases. This is apparent in the range response, which for a broadside target is the sum of the auto- and cross-correlations, as stated in (4.34).

The resulting range response when using an up- and a down-chirp as MIMO waveforms has the same range sidelobe structure near the peak as a single LFM, where the auto-correlation function dominates the MIMO range response. Instead of decaying to very low levels the contribution of the cross-correlation of the waveforms is apparent, even for relatively large delays.

From this analysis, we see that the simple case of an up- and down-chirp may provide acceptable peak sidelobe performance. Another figure of merit in waveform design is the integrated sidelobe level, which characterizes not simply the largest sidelobe but also includes the effect of all of the sidelobe energy. Clearly, the integrated sidelobe level is compromised by transmitting the second quasi-orthogonal LFM waveform.

This example of two waveforms was presented not as a recommendation for use in a MIMO radar but instead to present an example of the analysis that is required in choosing waveforms. We have seen previously that the zero-lag of the correlation matrix in (4.33),

namely, $\mathbf{R}_\phi(0)$, characterizes the antenna performance of a MIMO radar. Also, the structure of the matrix $\mathbf{R}_\phi(\tau)$ for $\tau \neq 0$ characterizes the range sidelobe performance, which is also critical to the operational utility of a radar system. The MIMO signal correlation matrix once again describes the capability of a set of waveforms to realize an effective MIMO radar.

4.6 | APPLICATIONS OF MIMO RADAR

Since the phased array is a special case of a MIMO radar, numerous such radar systems are currently operational. However, radars that attempt to transmit orthogonal waveforms are experimental at the time of this writing. Examples include a MIMO synthetic aperture radar (SAR) lab experiment described in [16] and a flight tested MIMO ground moving target indication (GTMI) system, which is documented in [17]. This section will demonstrate how the methods developed above may be applied to assess the performance of a MIMO radar when tasked with a particular mission.

4.6.1 MIMO SAR

A SAR system is able to form a high-resolution map of the earth's surface. Traditionally, a very large antenna was required to achieve desirable image resolutions, but SAR systems accomplish this with a reasonably sized antenna by transmitting pulses while moving the antenna that effectively sample the desired aperture [18]. This technique has been employed in numerous imaging radar systems since the 1950s, but it is clear that these systems may be improved by using MIMO radar techniques.

Phased array antennas have been used in SAR systems, primarily to provide enhanced flexibility with electronic scanning. The utility of these spatial degrees of freedom has been in the mitigation of range and Doppler ambiguities, which can severely impair image quality and limit area coverage rates. However, to realize these benefits, transmit gain must be sacrificed. Unlike the problem of detecting targets against a benign clutter background, SAR systems are often not limited by additive, thermal noise. Instead, performance is often driven by multiplicative noise sources [19]. As we shall see, mitigating multiplicative noise, such as from Doppler ambiguous clutter, introduces constraints on SAR systems that limit the quality or quantity of the generated imagery. Further, we will see how MIMO SAR can relax some of these requirements and improve the system performance.

To maximize area coverage rate, a SAR system may operate in a stripmap mode where it is able to continuously image as the sensor moves forward with a speed of v . A fundamental limit constrains the achievable cross-range resolution and area coverage rate of such a system. Let R_{swath} denote the depth of the range swath that is collected. The rate at which a map is formed is described by the area coverage rate, calculated by $ACR \triangleq vR_{swath}$.

The extent of the range swath is limited by the radar's ability to unambiguously measure range. Suppose that the radar transmits pulses with a pulse repetition frequency (PRF) of f_p (expressed in pulses per unit time). This limits the range swath to be $R_{swath} < (c/2)/f_p$, where c is the speed of waveform propagation.

Let $1/\delta x$ be the along-track sampling rate where

$$\delta x \triangleq v/f_p \quad (4.35)$$

is the distance that the platform travels between pulses. In many cases, along-track sampling requirements are the limiting factor for area coverage rate. This occurs when the time of flight of the pulse is large relative to the antenna size and the platform velocity. Also, along-track sampling may be limited by the throughput provided by the downlink, which prevents data from being collected beyond a certain rate. In these cases, the along-track sampling requirements impose an upper bound on achievable area coverage rate:

$$\text{ACR} < \delta x c/2 \quad (4.36)$$

SAR systems resolve targets in cross-range by exploiting the pulse-to-pulse phase variation across the imaged scene. If the along-track sampling rate is insufficient to unambiguously sample this Doppler frequency, returns from two very different cross-range locations will have the same phase progression. As a result, when an image is formed, undesired energy from these ambiguous returns will appear to be at the same location as the desired return. The impact of this on image quality is captured by computing the along-track ambiguity-to-signal ratio [20], which is a significant source of multiplicative noise in SAR imagery and can seriously degrade image contrast.

The along-track sampling rate must be sufficiently large to handle the Doppler bandwidth of the illuminated scene. Consequently, the impact of these ambiguities is a strong function of the radiation pattern of the antenna, which acts as a spatial filter. A larger antenna will have a narrower beam pattern that rejects returns with Doppler frequencies far from zero. If an antenna of length D is used for transmit and receive, then the along-track sampling rate used in practice typically corresponds to $\delta x \leq D/2$, though $\delta x = D/4$ is preferred.

Increasing the antenna size allows lower along-track sampling rates and thus provides increased area coverage (see (4.36)). However, it results in lower along-track resolution for stripmap (or a smaller scene size for spotlight SAR) because a larger antenna has a narrower beamwidth. For stripmap, this means that the target of interest is illuminated for a shorter amount of time limiting the integration angle and therefore limits resolution. (For spotlight, this results in a smaller image since a smaller patch on the ground is illuminated). The finest cross-range resolution, Δ_{CR} , is related to the antenna size by the inequality

$$\Delta_{CR} \geq \frac{D}{2} \quad (4.37)$$

In practice, this bound may be relaxed by applying a weighting to the antenna, which effectively broadens the mainlobe, however, it is still proportional to the antenna length, D .

We see that using a larger antenna allows us to use a lower PRF, thus achieving a higher area coverage rate but also resulting in coarser resolution. A standard approach to improve area coverage while preserving resolution is to employ a Vernier array [21]. In this configuration, a single radiating element is used on transmit and multiple elements are used on receive.

A Vernier array can be implemented by using a phased array that is spoiled on transmit. This array is divided into a number of subarrays. On transmit, a single subarray is used to illuminate a large area on the ground, but on receive N subarrays are used and are digitized to form N spatially diverse receive channels. If D is the length of a subarray, then the effective array length is ND . This array requires the along-track sampling rate corresponding to a subarray of length ND but can still provide the cross-range resolution of an antenna of length D . For a fixed velocity, by using N subarrays on receive, the radar

TABLE 4-2 ■ Properties of the virtual arrays corresponding to physical arrays with M transmitters and N receivers.

Configuration	Length	Spacing
Vernier array	$N \times D/2$	$D/2$
Dense MIMO	$N \times D/2$	$1/M \times D/2$
Sparse MIMO	$MN \times D/2$	$D/2$

can use a PRF that is decreased by a factor of N without impacting image quality and can realize a potential increase in area coverage rate.

The properties of the virtual arrays for these designs are given in Table 4-2. Note that the Vernier array is a special case of either a dense MIMO array or a sparse MIMO array. The results in the table for the MIMO configurations coincide with the Vernier array when $M = 1$.

The Dense MIMO Array Consider a receive array that consists of N receive elements that are of length D and spaced at an interval of D so that the receive array is contiguous. The SAR system uses M transmit elements that are spaced by D/M . The corresponding virtual array is sampled at an interval of $1/M \times D/2$. Note that the effective length of the dense MIMO array is the same as the corresponding Vernier array. However, a higher along-track sampling rate can be achieved by the dense MIMO array if it uses the same PRF as the Vernier array.

The Sparse MIMO Array Once again, begin with a Vernier array of N elements. Now, distribute the M transmitters such that there is a separation of ND between transmit subarrays. In this case, the spacing between virtual phase centers is $D/2$, as in the Vernier array case, but the resulting virtual array is M times as long as that of the Vernier array. This allows the sparse MIMO array to use a PRF that is M times lower than the Vernier array case, which provides a commensurate increase in area coverage rate.

Note that the sparse MIMO array is presented in the bottom of Figure 4-2. The dense MIMO array is similar to the the filled MIMO array, but the spacing between the transmit phase centers is such that they must fit between the first two receive phase centers.

An example of MIMO SAR is presented in [16]. An analysis of MIMO SAR collection approaches is found in [22]. These concepts for MIMO SAR can also be extended to synthetic aperture sonar (SAS) systems [23], which face tremendous challenges in maintaining sufficient along-track sampling rates with reasonable area coverage rates to the relatively slow speed at which sound propagates.

4.6.2 MIMO GMTI

By studying the angular PSF and the steered response, we saw that a MIMO radar has improved angular resolution and has the ability to re-steer its transmit beam digitally to enable a higher area coverage rate compared with the phased array. We will see that this provides enhanced detection of slow moving targets in the presence of strong clutter returns for ground moving target indication (GMTI) radar systems.

Recall that the enhanced capability afforded by transmitting orthogonal waveforms comes at the cost of SNR. However, in the case of GMTI, for example, this may be

recovered by dwelling longer. Just as the phased array uses a narrow pencil beam to scan a surveillance area, a MIMO radar is able to digitally emulate this. While the power observed on the ground at any given instant will be lower for the radar using orthogonal waveforms, it will cover a larger area. Consequently, a radar transmitting M orthogonal waveforms will have an SNR that is lower than a phased array by a factor of M , but the target may remain in the beam for M times as long since the transmit beam is larger. Unlike in SAR, where we incurred a penalty of M in terms of SNR, this may be recovered in GMTI, so long as the target returns remain coherent during the processing interval.

By examining the angle ambiguity function (or point spread function) for MIMO radar, we have demonstrated that a radar using orthogonal waveforms has superior angular resolution. In addition, with its ability to digitally resteer the transmit beam, it can dwell on a particular target for a longer period of time, thereby providing improved Doppler resolution while preserving area coverage rate. By transmitting orthogonal waveforms, a GMTI system is able to more effectively reject clutter and detect slow moving targets.

Note that this improved Doppler resolution could also be achieved by spoiling on transmit. However, recall that the angular resolution of the phased array is not improved by spoiling.

The key figure of merit in GMTI is signal-to-interference plus noise ratio (SINR) loss [24]. For a target at a particular angle and velocity (more precisely, a target with a space-time steering vector \mathbf{s}), the SINR loss measures the drop in SNR as a result of ground clutter. It is defined as the ratio of SINR (target vs. clutter-plus-noise) to SNR (target vs. noise). If the clutter-plus-noise space-time covariance matrix is \mathbf{R} and the noise-only space-time covariance matrix is \mathbf{R}_N , then the SINR loss for a target with space-time steering vector \mathbf{s} is given by

$$\text{SINR} = \frac{\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}}{\mathbf{s}^H \mathbf{R}_N^{-1} \mathbf{s}} \quad (4.38)$$

In Figure 4-10, a single angle bin is considered. Since the radar using orthogonal waveforms has better angular resolution than the phased array, the orthogonal waveforms angle bin contains a smaller angular extent. As a result, clutter is present at fewer Doppler frequencies within this angle bin. This allows the detection of slow moving targets and an improvement of minimum detectable velocity (MDV).

4.6.3 Distributed Apertures

A phased array uses a set of radiating elements to transmit a high gain beam in a desired direction, which obviously requires a high degree of synchronization among the transmit elements and a well calibrated system. There has been interest in extending this functionality to a collection of platforms that may be distributed over some area. A key challenge using such a distributed aperture to beamform on transmit is maintaining the required coherence from platform to platform.

One approach to solving this problem can be considered in a MIMO context. We have seen how, by transmitting orthogonal waveforms, a MIMO radar can digitally resteer its transmit beam. If the waveforms are orthogonal, then transmit degrees-of-freedom are preserved for manipulation after reception. An example of this approach as applied to a constellation of radar satellites is presented in [25].

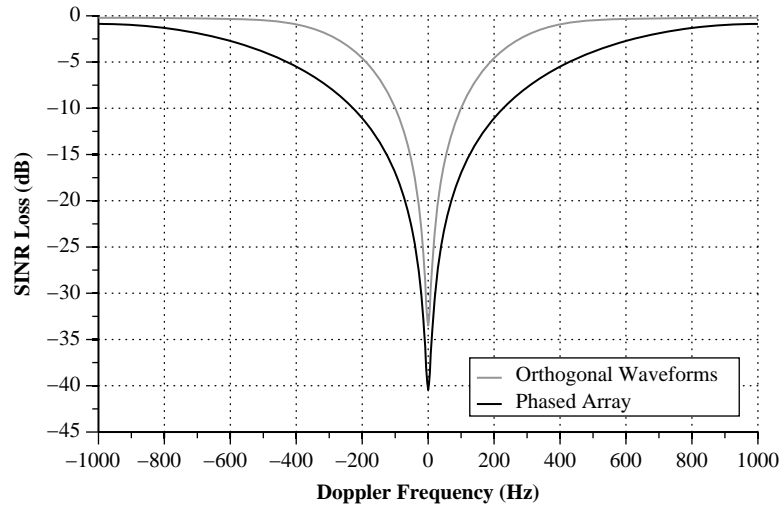


FIGURE 4-10 ■ Notional SINR loss for GMTI. If SINR loss is near 0 dB, then a target with the velocity corresponding to that Doppler frequency is easily detected in the presence of clutter. The SINR loss for orthogonal waveforms is significantly lower than that for the conventional phased array for many frequencies, so an improved minimum detectable velocity is expected.

4.6.4 Beampattern Synthesis

Phased array radars that employ digital beamforming on receive are able to spoil their beam on transmit to cover a larger area. In doing so, the search rate may be increased at the cost of signal-to-noise ratio. While this is typically accomplished by applying a phase taper across the array, this can also be considered in a MIMO context. As described in [26], beam spoiling by transmitting orthogonal waveforms relaxes some of the requirements on the radar hardware when clutter is limiting detection performance.

As described before, the transmit beampatterns that may be synthesized are a function of the MIMO signal correlation matrix, \mathbf{R}_ϕ . To synthesize a desired pattern, the appropriate correlation matrix must be identified. Further, a suite of signals must be found that possesses these correlations. Some examples of such techniques are presented in [27,28].

4.7 | SUMMARY

As is hopefully evident from this chapter, analysis of the utility of MIMO techniques for a particular radar application may not be straightforward. It should also be clear that transmitting orthogonal waveforms is not advantageous in every situation. The primary goal of this chapter has been to provide a framework for evaluating the appropriateness of a particular suite of MIMO waveforms for a specific radar mission. This is necessary to decide if performance will be enhanced by using a MIMO radar instead of a traditional phased array configuration.

By synthesizing virtual phase centers, we saw how a MIMO radar can provide enhanced angular resolution. This may be of utility in the case of GMTI and air-to-air radar. A similar analysis revealed the ability of a MIMO radar to trade peak gain for a more flexible transmit pattern. An application to SAR was presented.

The performance of a MIMO radar was quantified by examining three standard quantities in array processing: the steered response, the beampattern, and the (angular) point spread function. By evaluating these patterns, the ability of an antenna to perform its competing roles of enhancing gain in some directions while rejecting signals for others while providing a large access area is determined.

A key result was to demonstrate that the performance of a MIMO radar system is characterized by its transmitted signal correlation matrix, \mathbf{R}_ϕ . This allows us to quantify the impact on performance when truly orthogonal waveforms are not available. It also allows the traditional phased array architecture to be analyzed as a subset of MIMO radar. This insight is comforting: if an analysis is conducted to optimize \mathbf{R}_ϕ , a possible outcome is that the phased array is proven to be the optimal configuration in some cases; in other cases, as we have seen, transmitting orthogonal waveforms provides tremendous performance benefits. Since all radars are MIMO radars, the question is not whether to use a phased array or a MIMO radar but rather what signal correlation matrix is optimal?

4.8 | FURTHER READING

Just like MIMO communications before it, MIMO radar has received much attention in the research literature. A comprehensive summary of current research into MIMO radar can be found in the book edited by Li and Stoica [29]. For a review of MIMO with colocated antennas, see [30].

4.9 | REFERENCES

- [1] Bliss, D.W. and Forsythe, K.W., MIMO Radar and Imaging: Degrees of Freedom and Resolution, in *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 54–59, November 2003.
- [2] Silver, S., *Microwave Antenna Theory and Design, Radiation Laboratory Series*, McGraw-Hill Book Company, Inc., New York, 1949.
- [3] Dorey, J. and Garnier, G., RIAS, Radar à impulsion et antenna synthétique, *L'Onde Electrique*, vol. 69, no. 6, pp. 36–44, 1989.
- [4] Hoor, R.T. and Kassam, S.A., The Unifying Role of the Coarray in Aperture Synthesis for Coherent and Incoherent Imaging, *Proceedings of the IEEE*, vol. 78, no. 4, pp. 735–752, April 1990.
- [5] Lockwood, G.R., Li, P.-C., O'Donnell, M., and Foster, F.S., Optimizing the Radiation Pattern of Sparse Periodic Linear Arrays, *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 43, no. 1, pp. 7–14, January 1996.
- [6] Fishler, E., Haimovich, A., Blum, Jr., R.S., Cimini, L.J., Chizhik, D., and Valenzuela, R.A., Spatial Diversity in Radars-Models and Detection Performance, *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 823–838, March 2006.
- [7] Chernyak, V.S., *Fundamentals of Multisite Radar Systems*, CRC Press, Boca Raton, FL, 1998.
- [8] Haimovich, A.M., Blum, R.S., and Cimini, L.J., MIMO Radar with Widely Separated Antennas, *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 116–129, January 2008.
- [9] Forsythe, K.W. and Bliss, D.W., MIMO Radar Waveform Constraints for GMTI, *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 21–32, February 2010.

- [10] H. V. Henderson and S. R. Searle. The vec-Permutation Matrix, the vec Operator, and Kronecker Products: A Review *Linear and Multilinear Algebra*, 9:271–288, 1981.
- [11] R.A. Monzingo and T.W. Miller. *Introduction to Adaptive Arrays*. SciTech, Raleigh, NC, 2011.
- [12] S.L. Campbell and C.D. Meyer. *Generalized Inverses of Linear Transformations*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2009.
- [13] Bekkerman, I. and Tabrikian, J., Target Detection and Localization Using MIMO Radars and Sonars, *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3873–3883, October 2006.
- [14] Johnson, D.H. and Dudgeon, D.E., *Array Signal Processing: Concepts and Techniques*, Prentice-Hall, Upper Saddle River, NJ, 1993.
- [15] San Antonio, G., Fuhrmann, D.R., and Robey, F.C., MIMO Radar Ambiguity Functions, *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 1, pp. 167–177, June 2007.
- [16] Rennich, P.K., Four-Platform Distributed MIMO Radar Measurements and Imagery, paper presented at the IEEE Radar Conference, May 2009.
- [17] Kantor, J. and Davis, S.K., Airborne GMTI Using MIMO Techniques, in *IEEE Radar Conference*, pp. 1344–1349, May 2010.
- [18] J. C. Curlander and R. N. McDonough. *Synthetic Aperture Radar: Systems and Signal Processing*. Wiley, New York, 1991.
- [19] Carrara, W.G., Goodman, R.S., and Majewski, R.M., *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*, Artech House, Boston, MA, 1995.
- [20] Mehliis, J.G., Synthetic Aperture Radar Range-Azimuth Ambiguity Design and Constraints, in *IEEE International Radar Conference*, pp. 143–152, 1980.
- [21] Kock, W.E., Extending the Maximum Range of Synthetic Aperture (Hologram) Systems, *Proceedings of the IEEE*, vol. 60, no. 11, pp. 1459–1460, November 1972.
- [22] B. Correll. Efficient Spotlight SAR MIMO Linear Collection Configurations. *IEEE Journal of Selected Topics in Signal Processing*, 4(1):33–39, February 2010.
- [23] Davis, M. and Cook, D., MIMOSAS: Improving SAS Performance with Transmit Diversity, in *Proceedings of the 4th Underwater Acoustic Measurements Conference*, June 2011.
- [24] W.L. Melvin. A STAP Overview. *IEEE Aerospace and Electronic Systems Magazine*, 19(1): 19–35, January 2004.
- [25] Steyskal, H., Schindler, J.K., Franchi, P., and Mailloux, R.J., Pattern Synthesis for TechSat21—A Distributed Space-Based Radar System, *IEEE Antennas and Propagation Magazine*, vol. 45, no. 4, pp. 19–25, August 2003.
- [26] Rabideau, D.J. and Parker, P., Ubiquitous MIMO Multifunction Digital Array Radar, in *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 1057–1064, November 2003.
- [27] Stoica, P., Li, J., and Xie, Y., On Probing Signal Design for MIMO Radar, *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4151–4161, August 2007.
- [28] Fuhrmann, D.R. and San Antonio, G., Transmit Beamforming for MIMO Radar Systems Using Signal Cross-Correlation, *IEEE Transactions on Aerospace and Electronic Systems*, 44, no. 1, pp. 171–186, January 2008.
- [29] Li, J. and Stoica, P., *MIMO Radar Signal Processing*, Wiley-IEEE Press, New York, 2008.
- [30] Li, J. and Stoica, P., MIMO Radar with Colocated Antennas, *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, September 2007.

4.10 | PROBLEMS

1. What distinguishes a MIMO radar from a phased array radar?
2. True or false: A radar transmitting orthogonal waveforms will observe the same SNR for a particular target as a phased array radar using the same integration time.
3. Show that the optimal spatial weights in (4.26) satisfy the normal equations, given by (4.25). The following facts may be useful:
 - The pseudoinverse is at least a weak inverse: $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$
 - The pseudoinverse of a transposed matrix is the transpose of its pseudoinverse: $(\mathbf{A}^T)^+ = (\mathbf{A}^+)^T$
 - The matrix product of two Kronecker products is the Kronecker product of the matrix products: $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$ for suitably sized matrices.
4. Consider a set of M signals, $s_1(t), \dots, s_M(t)$, defined by

$$s_m(t) \triangleq \begin{cases} \frac{1}{\sqrt{T}} \exp\{i2\pi \frac{m}{T}t\}, & \text{for } t \in [0, T) \\ 0, & \text{otherwise} \end{cases}$$

for $m = 1, \dots, M$. Using the definition given in (4.2), show that these signals are mutually orthogonal.

5. Duplicate the gain patterns shown in Figure 4-6. Assume that the transmit and receive arrays are uniform linear arrays. For the elements of the steering vector, $\mathbf{s}(\theta)$, use

$$s_m(\theta) = e^{i2\pi \sin(\theta)x_m}$$

where x_m is the location of element m along the x -axis, in units of wavelengths.

Radar Applications of Sparse Reconstruction and Compressed Sensing

Jason T. Parker, Matthew A. Ferrara, Lee C. Potter

Chapter Outline

5.1	Introduction	147
5.2	CS Theory	150
5.3	SR Algorithms	166
5.4	Sample Radar Applications	183
5.5	Summary	196
5.6	Further Reading	196
5.7	Acknowledgments	197
5.8	References	197
5.9	Problems	207

5.1 INTRODUCTION

Sparse reconstruction and design through randomization have played significant roles in the history of radar signal processing. A recent series of theoretical and algorithmic results known as compressive or compressed sensing (CS) has ignited renewed interest in applying these ideas to radar problems. A flurry of research has explored the application of CS approaches as well as closely related sparse reconstruction (SR) techniques to a wide range of radar problems. This chapter will provide some historical context for CS, describe the existing theoretical results and current research directions, highlight several key algorithms that have emerged from these investigations, and offer a few examples of the application of these ideas to radar.

5.1.1 Organization

The chapter is organized into three sections. Section 5.2 develops the motivation and theoretical framework for SR and CS. We attempt to motivate these ideas from a radar perspective while also highlighting intuitions and connections with basic linear algebra and optimization theory. Section 5.3 explores the myriad of available algorithms for solving SR problems and the provable performance guarantees associated with these algorithms when used in conjunction with measurements which satisfy CS design criteria. Section 5.4 concludes with a selection of examples that illustrate how to apply these ideas to radar problems.

5.1.2 Key Points

- Many interesting radar problems in imaging, detection, tracking, and identification can be formulated in a linear model framework that is amenable to SR and CS techniques. In fact, the linear model framework is not limited to point scatterer assumptions, free-space propagation, or weak scattering scenes.
- Radar signals often exhibit sparsity or compressibility in a known basis, such as the basis of point scatterers for high frequency synthetic aperture radar (SAR) data.
- Many radar problems are underdetermined and must be regularized with additional information to obtain unique solutions. Sparsity in a known basis represents one appealing candidate but leads to intractable NP-hard optimization problems.
- A key notion in CS theory is to overcome the combinatorial complexity of seeking sparse solutions through convex relaxation with the ℓ_1 norm.
- SR by itself is not CS. CS involves combining SR algorithms with constraints on the measurement process, typically satisfied through randomization, which leads to provable performance guarantees.
- CS performance guarantees are predicated on conditions on the forward operator like the restricted isometry property or mutual coherence.
- A wide range of SR algorithms are available.
- Penalized least squares techniques directly solve a convex (or nonconvex) relaxation of the ℓ_0 optimization problem and offer excellent performance coupled with relatively high computational burden.
- Iterative thresholding methods offer simplified algorithms with performance guarantees, but in some cases these guarantees are weaker than those for penalized least squares.
- Greedy algorithms offer faster heuristic approaches to SR problems that provide limited performance guarantees.
- New trends in CS include approaches inspired by Bayesian and information theoretic approaches.
- Many CS guarantees are sufficient, conservative guarantees. SR algorithms may produce desirable results for radar problems even when the conditions for these guarantees are not met. Phase transition plots offer a concise method for representing typical numerical algorithm performance on problems of interest.
- By incorporating structured sparsity information, CS performance can be further enhanced.

5.1.3 Notation

Variables used in this chapter include the following:

\mathbf{x}^{true} = true value of the unknown vector

\mathbf{A} = the forward operator relating unknowns to measurements

\mathbf{y} = the vector of collected measurements

\mathbf{e} = the additive disturbance signal

N = number of elements in the unknown vector

M = number of measurements collected

$s = \|\mathbf{x}^{true}\|_0$, the sparsity of the unknown vector

$\delta = M/N$, the undersampling ratio
 $\rho = s/M$, the normalized sparsity
 σ = the energy bound on the additive disturbance signal \mathbf{e}
 f_0 = radar center frequency (Hz)
 λ = wavelength (meters)
 $\omega_0 = 2\pi f_0$, center frequency in (radians per second)
 c = speed of light (meters per second)
 d = array inter element spacing (meters)
 $f = c/\lambda$, frequency (Hz)
 $k = 2\pi/\lambda$, wavenumber (radians per meter)
 J = number of spatial channels
 K = number of slow-time pulses
 L = number of range gates in fast time
 T = slow-time sampling period (seconds).

5.1.4 Acronyms

Acronyms used in this chapter include the following:

AMP	approximate message passing
BP	basis pursuit
BPDN	basis pursuit denoising
CoSaMP	compressive sampling matching pursuits
CPI	coherent processing interval
CS	compressed sensing
DFT	discrete Fourier transform
DLS	data least squares
DWT	discrete wavelet transform
EM	expectation maximization
FISTA	fast iterative shrinkage-thresholding algorithm
FOCUSS	focal undetermined system solver
IHT	iterative hard thresholding
IRLS	iterative reweighted least squares
ISTA	iterative shrinkage-thresholding algorithm
LaMP	lattice matching pursuit
LASSO	least absolute shrinkage and selection operator
MAP	<i>maximum a posteriori</i>
MIMO	multiple input multiple output
ML	maximum likelihood
MM	majorization minimization
MMSE	minimum mean squared error
MP	matching pursuits
MTI	moving target indication
NESTA	Nesterov's algorithm
OMP	orthogonal matching pursuits
PSF	point spread function
RIC	restricted isometry constant
RIP	restricted isometry property
SAR	synthetic aperture radar

SP	subspace pursuit
SR	sparse reconstruction
STAP	space time adaptive processing
SVD	singular value decomposition
TV	total variation
ULA	uniform linear array.

5.2 | CS THEORY

The origin of the name compressed sensing lies in a particular interpretation of CS algorithms as an approach to signal compression. Many systems sample a signal of interest at a rate above the Nyquist sampling rate dictated by the signal's bandwidth. This sampled signal is then transformed to a basis where a few large coefficients contain most of the signal's energy. JPEG2000 is an excellent example of this sort of processing, relying on a wavelet transformation. The signal can then be compressed by encoding only these large signal coefficients and their locations.

Since the signal can be encoded with just a few coefficients, it seems natural to ask if the relatively large number of measurements is required in the first place. The original sampling rate was dictated by Nyquist sampling theory to guarantee the preservation of an arbitrary band-limited signal. However, perhaps one can use the knowledge that the signal will be represented by only a few nonzero components in a known basis, such as a wavelet transform, to reduce the required data acquisition. It turns out that, under certain conditions, a relatively small number of randomized or specially designed measurements of the signal can be used to reconstruct this sparse representation. The key is that we do not need to know which coefficients are nonzero; we require knowledge only of the basis or dictionary from which these elements or atoms are drawn. In fact, in the case of noiseless signals, this reconstruction from a reduced data set will actually be perfect! Furthermore, we shall see that the reconstruction of the signal will be well-behaved both in the presence of noise and when the signal is only approximately sparse. Because a reduced data set is being collected and compression is accomplished through the sampling procedure itself, this process is termed compressed sensing.

This combination of randomized measurements with a sparse representation forms the heart of CS. Indeed, CS combines measurement randomization with SR to provide performance guarantees for solving ill-posed linear inverse problems [1,2]. We will explore the implications and interpretation of this statement at length throughout this chapter. First, we will define the problem of interest and explore its relevance to radar.

5.2.1 The Linear Model

Many radar signal processing problems can be represented with a linear measurement model. In particular, consider an unknown complex-valued signal of interest $\mathbf{x}^{true} \in \mathbb{C}^N$. We collect a set of measurements of the signal $\mathbf{y} \in \mathbb{C}^M$ using a forward model or measurement operator $\mathbf{A} \in \mathbb{C}^{M \times N}$ with additive noise $\mathbf{e} \in \mathbb{C}^M$, that is,

$$\mathbf{y} = \mathbf{A}\mathbf{x}^{true} + \mathbf{e} \quad (5.1)$$

Our fundamental goal will be to solve the inverse problem of determining the vector \mathbf{x}^{true} from the noisy measurements \mathbf{y} . As we will see, even in the noise-free case $\mathbf{e} = \mathbf{0}$, this problem is nontrivial with multiple feasible solutions.

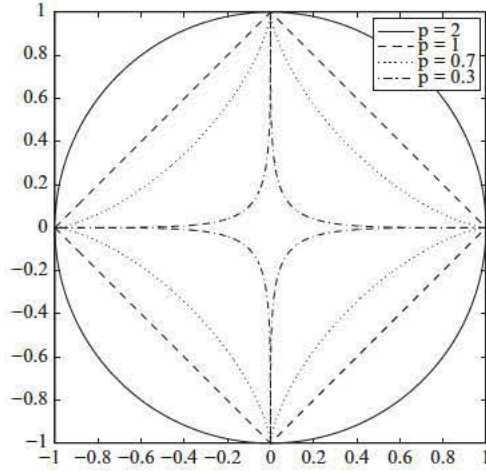


FIGURE 5-1 ■ Unit balls, that is, $\{x : \|x\|_p = 1\}$, for various values of p . Notice that the unit balls are convex for $p \geq 1$.

We define the ℓ_p norm of a vector as $\|x\|_p = (\sum |x_i|^p)^{1/p}$. For $p = 2$, the ℓ_p norm is the traditional notion of Euclidean distance and represents the energy in a signal. We shall see that the ℓ_1 norm, sometimes referred to as the taxicab norm since it is the distance one must travel between two locations on a grid of roads, plays a special role in CS. We also define, through a mild abuse of notation, the ℓ_0 norm to be the number of nonzero components in the vector.¹ We similarly define $\|x\|_\infty$ as the magnitude of the largest coefficient in x . Figure 5-1 depicts the unit balls (i.e., the set of points with norm one) in two dimensions for several values of p .

We will denote the sparsity, or number of nonzero coefficients, of the unknown signal as $s = \|x^{true}\|_0$. We will also assume that the energy of the additive noise term is bounded as $\|e\|_2 \leq \sigma$. In addition, it is convenient to assume that each column of A has unit ℓ_2 norm. This assumption does not sacrifice any generality because a simple rescaling of x can be used to normalize the columns if needed.

The problems of interest are typically ill-posed in the sense that $M < N$. Put another way, there are fewer equations than unknowns, and the solution is not unique. We shall return to this topic in detail in Section 5.2.3.

For now, we define the undersampling ratio $\delta = M/N$ to quantify the severity of the undetermined nature of the operator A . As $\delta \rightarrow 1$ the problem becomes fundamentally easier to solve because more measurements are available per unknown. Similarly, we define the normalized sparsity as $\rho = s/M$. As $\rho \rightarrow 1$ the problem becomes fundamentally more difficult to solve since we need to recover more nonzero coefficient values for the same number of measurements. A sharp change in algorithm performance in terms of δ and ρ is referred to as a phase transition and will be discussed in more depth in Section 5.3.5.2.

5.2.2 The Linear Model in Radar

The linear model defined in (5.1) can be used to describe a wide range of radar signal processing applications. Indeed, any radar data set that obeys an approximately linear

¹This name is selected since $\lim_{p \rightarrow 0} \|x\|_p = \|x\|_0$. However, the ℓ_0 norm is not a norm, since $\|\alpha x\|_0 = \|x\|_0$ for any scalar $\alpha \neq 0$. In fact, ℓ_p ceases to be a norm for $p < 1$ because it violates the triangle inequality. For this reason ℓ_p with $p < 1$ is referred to as a pseudo-norm.



FIGURE 5-2 ■ A notional depiction of a collection geometry for RF tomography. Each of the antennas transmits RF signals whose echoes are received and processed by all of the nodes. Some form of coding or time/frequency multiplexing is used to make the signals distinguishable at the receivers. This separability is depicted here with different coloring on the transmitted waveforms.

relationship between the signal of interest and the measured data can be captured in this framework. One key assumption is that multiple scattering by the target objects can be neglected, a common assumption referred to as the Born approximation² [4].

Intuitively, one can consider the matrix \mathbf{A} as a dictionary whose k -th column, or atom, represents the data that would be collected if only a single element of \mathbf{x}^{true} were nonzero. Our goal of inferring \mathbf{x}^{true} from \mathbf{y} can thus be viewed as the task of finding the dictionary entries that best describe the collected data. To make this idea concrete, we will briefly sketch two examples of expressing radar data in this framework. The reader is referred to the literature or earlier chapters in this series for more detailed radar data models.

5.2.2.1 Radio Frequency Tomography Example

We will now consider a frequency-domain example using several spatially distributed narrowband sensors. A radio frequency (RF) tomographic system illuminates a scene with several spatially distributed radar transmitters. The echoes from these sensors are received by a constellation of radar receivers for subsequent processing. Figure 5-2 depicts a notional scene with a network of these distributed sensors. Even traditional monostatic SAR can be easily viewed in a tomographic framework (see e.g., [5]).

Data from a set of M arbitrary-geometry bistatic scattering experiments will be processed jointly to image a scene. Each scattering experiment will use a receiver located

²Another method of model linearization is the Rytov approximation [3], which linearizes the nonlinear equation satisfied by the phase of the scalar field. For most radar applications the Born approximation is appropriate, and it is by far the most commonly used model. Thus, we will limit our discussion of linear approximations to the single-scattering setting. As we will show in Section 5.4.4, this will not preclude the consideration of multiple-scattering scenarios.

at position $\mathbf{r}_m \in \mathbb{R}^3$ and a transmitter at $\mathbf{t}_m \in \mathbb{R}^3$ operating at wavenumber $k_m = \omega_m/c = 2\pi f_m/c$, where c is the speed of propagation. The RF transmission frequency is given in hertz as f_m . Data in this format can of course be synthesized from a pulsed system by match filtering and transforming the processed signal to the frequency domain. One can think of these data as phase history data for a multistatic SAR collection after dechirp or deramp processing [5,6].

Consider an isotropic point scatterer located at an arbitrary position $\mathbf{q} \in \mathbb{R}^3$. Assuming scalar wave propagation governed by the Helmholtz equation, the Green's function (section 7.3 of [7]) describing the spherical wave emanating from a point source \mathbf{t}_m and impinging on the point \mathbf{q} is given by

$$G(\mathbf{t}_m, \mathbf{q}, k_m) = \frac{e^{jk_m \|\mathbf{t}_m - \mathbf{q}\|_2}}{4\pi \|\mathbf{t}_m - \mathbf{q}\|_2} \quad (5.2)$$

To obtain the field received by a point source with an arbitrary transmit waveform $a_m(k) = \int_{-\infty}^{\infty} \hat{a}_m(t) e^{jkt} dt$, the frequency-domain Green's function $G(\mathbf{t}_m, \mathbf{q}, k_m)$ is modulated by the complex-valued frequency-domain waveform³ $a_m(k_m)$; thus the field received at point \mathbf{q} is $a_m(k_m)G(\mathbf{t}_m, \mathbf{q}, k_m)$. Similarly, if the scattering potential at \mathbf{q} is given by $x(\mathbf{q})$, then the measured field for scattering experiment m from the single point target would be

$$y_m = x(\mathbf{q})a_m(k_m)G(\mathbf{t}_m, \mathbf{q}, k_m)G(\mathbf{q}, \mathbf{r}_m, k_m) \quad (5.3)$$

The antenna patterns for the transmit and receive antennas are omitted for simplicity. If we linearize the scalar data model using the Born approximation [9] and discretize the scene into N voxels with locations \mathbf{q}_n and reflectivities $\mathbf{x} = [x(\mathbf{q}_1)x(\mathbf{q}_2)\dots x(\mathbf{q}_N)]^T$, then we can obtain the total measured field for experiment m as

$$y_m = \sum_{n=1}^N x(\mathbf{q}_n)a_m(k_m)G(\mathbf{t}_m, \mathbf{q}_n, k_m)G(\mathbf{q}_n, \mathbf{r}_m, k_m) \quad (5.4)$$

We can express the complete data collection in the form of (5.1) by defining $\mathbf{A} \in \mathbb{C}^{M \times N}$ with $A_{mn} = a_m(k_m)G(\mathbf{t}_m, \mathbf{q}_n, k_m)G(\mathbf{q}_n, \mathbf{r}_m, k_m)$ and denoting the additive measurement noise as \mathbf{e} .

The inner products of the columns of \mathbf{A} can be related to values of the familiar point spread function (PSF) from SAR imaging [10]. Under the far-field assumption $\|\mathbf{t}\|_2 \gg \|\mathbf{q}\|_2$, the PSF reduces to the standard form encountered in parallel-ray computed tomography [6,11]. An example of this development in the time domain with arbitrary waveforms can be found in [12].

5.2.2.2 The Ambiguity Function

We next present two moving target indication (MTI) examples. The first example models the data for a single pulse using a single radar channel to discriminate delay and Doppler. The two-dimensional function describing the coupling between delay and Doppler for a radar waveform is known as the ambiguity function.

Consider the case of a radar with narrow fractional bandwidth and monostatic operation, that is, co-located transmitter and receiver antennas. Working in the time domain,

³This is because Green's function is a fundamental solution [8] to the scalar wave equation $(\nabla^2 + k^2)G = -\delta$, where δ is the Dirac delta function.

consider a complex baseband pulse, $u(t)$, modulated in quadrature by a carrier with frequency, $\omega_0 = 2\pi f_0$, to yield the transmit waveform, $p(t) = \text{Re}\{u(t)e^{j\omega_0 t}\}$, where $\text{Re}\{\}$ denotes the real operator. The echo of the transmitted pulse waveform encodes backscatter energy from the illuminated scene. Assume the illuminated scene consists of scatterers at range r with radial velocity, v . We can parametrize the complex scene reflectivity in terms of delay and Doppler as $x(\tau, \omega)$, where $\tau(r) = \frac{2r}{c}$ is the round-trip propagation time, and $\omega(v) = \frac{2\omega_0 v}{c}$ is the Doppler shift. The total received signal backscattered from the scene after quadrature demodulation is the baseband signal

$$y_B(t) = \int \int x(\tau, \omega) u(t - \tau) e^{-j\omega t} d\tau d\omega + n_B(t) \quad (5.5)$$

where the constant phase terms have been absorbed into the reflectivity, and $n_B(t)$ represents the circular white complex Gaussian baseband noise.

Our goal is to determine the reflectivity function $x(\tau, \omega)$ from the received data $y_B(t)$. This model can be discretized to obtain a set of equations in the form (5.1). If we discretize the scene reflectivity function x over delay and Doppler on a grid of points, $\{\tau_m, \omega_m\}$, to produce the vector \mathbf{x} and sample the received baseband signal y_B and noise $n_B(t)$ at times $\{t_m\}$ to obtain \mathbf{y} and \mathbf{e} , we obtain the linear system of equations

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \quad (5.6)$$

Each column \mathbf{a}_i of \mathbf{A} thus represents the received waveform for a scatterer with given Doppler and range (i.e., delay), and all columns share the same ℓ_2 norm, under the far-field assumption.

One simple approach to locating targets in the delay-Doppler plane is to derive a test statistic for a single hypothesized target at a particular delay-Doppler combination. We will pursue this approach briefly to highlight a connection between CS theory and traditional radar practice. Recall that the likelihood ratio test statistic for the existence of a single target with delay τ' and Doppler frequency ω' is the matched filter output [13],

$$\begin{aligned} \chi(\tau', \omega') &= \int y_B(t) u^*(t - \tau') e^{j\omega' t} dt \\ &= \int \int x(\tau, \omega) \mathcal{A}(\tau - \tau', \omega - \omega') d\tau d\omega + \int n_B(t) u^*(t - \tau') e^{j\omega' t} dt \end{aligned} \quad (5.7)$$

where the radar ambiguity function $\mathcal{A}(\tau, \omega)$ is given by $\mathcal{A}(\tau, \omega) = \int u(t) u^*(t - \tau) e^{j\omega t} dt$. Thus, the output of the matched filter is the convolution of the reflectivity $f(\tau, \omega)$ with the radar ambiguity function $\mathcal{A}(\tau, \omega)$, plus a filtered copy of the baseband noise. The shape of the ambiguity function can be adjusted by varying the pulse waveform $u(t)$. However, shaping of the ambiguity function is subject to a total volume constraint:

$$\int \int |\mathcal{A}(\tau, \omega)|^2 d\tau d\omega = \|u(t)\|_2^2 \quad (5.8)$$

Thus, while the waveform can be designed to reshape the ambiguity function to satisfy design criteria, improving the ambiguity response in one region will necessarily push energy into other parts of the delay-Doppler plane. For example, making the peak of the response narrower will typically raise the sidelobe levels. The sampled matched filter

outputs can be written

$$\chi = \mathbf{A}^H \mathbf{y} \quad (5.9)$$

The inner products of the columns of \mathbf{A} , denoted by \mathbf{a}_i , are samples of the ambiguity function: $|\mathbf{a}_i^H \mathbf{a}_j| = |\mathcal{A}(\tau_i - \tau_j, \omega_i - \omega_j)|$. This relationship is a crucial point in understanding the connection between CS and radar signal processing, as we will see in Section 5.2.5.

5.2.2.3 Multichannel Example

Now consider processing multiple pulses with a multichannel phased array radar. In contrast to the previous example, assume that matched filtering has already been performed on the individual pulses in fast time and focus on modeling the target response in the slow-time⁴ and channel dimensions. This phased array will transmit a series of pulses steered to a region of interest on the ground.⁵ The echoes from these pulses will be received on multiple channels connected to subarrays of the antenna. By coherently processing these returns, range, velocity, and angular bearing information can be extracted about moving targets. An MTI system treats nonmoving objects as undesirable clutter and attempts to suppress these returns using techniques like space-time adaptive processing (STAP) [15]. Figure 5-3 depicts a notional MTI scenario.

To be specific, consider a monostatic uniform linear array (ULA) consisting of J channels spaced equally at d meters apart. A coherent processing interval (CPI) for this system consists of data collected over K slow-time pulses with a sampling period of T seconds and L fast time range bins. We shall assume that the system is narrowband (i.e., the bandwidth $B \ll f_0$), where $f_0 = \frac{c}{\lambda}$ is the center frequency,⁶ and that pulse compression has already been performed. In addition, motion during a given pulse will be neglected.⁷

We will consider the data collected for a single range gate. The spatial-channel samples for pulse k will be denoted as a vector $\mathbf{y}_k \in \mathbb{C}^J$, while the complete space-time snapshot will be denoted $\mathbf{y} \in \mathbb{C}^{JK}$, where the data have been pulse-wise concatenated, that is,

$$\mathbf{y} = [\mathbf{y}_1^T \quad \mathbf{y}_2^T \quad \dots \quad \mathbf{y}_K^T]^T$$

For a CPI, we thus collect $M = JK$ measurements of the unknown scene.

⁴Slow time refers to the relatively long intervals between successive pulses from a coherent radar. Fast time refers to the time scale at which the electromagnetic pulse travels across the scene of interest, which is the same as range up to a scale factor when propagation is in a homogenous medium, such as free space.

⁵Our development here assumes that all the elements in the phased array transmit the same waveform during a given pulse, with the exception of phase shifts to steer the beam in the desired illumination direction. A more general approach involves using distinct waveforms on each of the transmit channels. This multiple-input multiple-output (MIMO) approach has recently received significant attention; see for example [14].

⁶This assumption allows time delays to be well approximated as phase shifts, which creates a correspondence between target velocity and the output bins of a fast Fourier transform (FFT) with respect to slow-time.

⁷This so-called stop-and-hop approximation is very reasonable for the short pulses associated with MTI platforms [16].

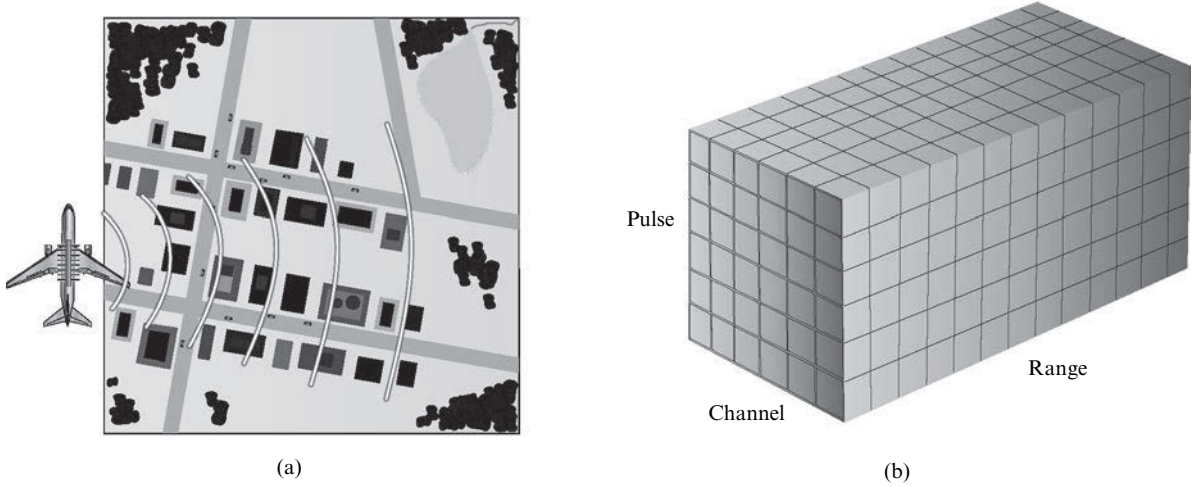


FIGURE 5-3 ■ (a) A simple depiction of an MTI data collection scenario. The targets of interest are the moving vehicles on the road network. (b) A notional rendering of the data cube collected during a CPI. Each element in the cube is a single complex-valued sample. The three dimensions are channels, pulses, and range. Similar diagrams are used to describe MTI data sets, along with much more detail, in [15,16].

At a given range, the response of the array over the CPI to a point target can be characterized with a space-time steering vector. First, consider a target response arriving at elevation angle θ and azimuth angle ϕ as measured by J spatial channels. Since the array is linear in the azimuth plane, there is a conical ambiguity in the arrival direction of a given signal characterized by the cone angle $\theta_c = \cos^{-1}(\cos \theta \sin \phi)$. The spatial frequency observed by the array for a given cone angle is then

$$f_s = \frac{d}{\lambda} \cos \theta_c$$

The spatial steering vector for a given cone angle is then $\mathbf{a}_s \in \mathbb{C}^J$, given by

$$\mathbf{a}_s(f_s) = [1 \quad \exp(j2\pi f_s) \quad \dots \quad \exp(j2\pi(J-1)f_s)]^T$$

where we have selected the first element as the zero-phase reference for the array. Similarly, we can define the normalized Doppler frequency as

$$f_d = \frac{2vT}{\lambda}$$

where v is the velocity of the target. The temporal steering vector describing the response of a single element across K time samples to a target at normalized Doppler f_d is then given as the length K vector

$$\mathbf{a}_t(f_d) = [1 \quad \exp(j2\pi f_d) \quad \dots \quad \exp(j2\pi(K-1)f_d)]^T$$

The combined space-time steering vector for a target is then given as the Kronecker product of the temporal and spatial steering vectors, that is,

$$\mathbf{a}(f_s, f_d) = \mathbf{a}_t(f_d) \otimes \mathbf{a}_s(f_s)$$

For a given range bin, the vector $\mathbf{a}(f_s, f_d)$ represents the data that the radar would collect over a CPI if only a single target having unit scattering amplitude were present at the angle-Doppler location encoded by f_s and f_d . Specifically, the steering vector $\mathbf{a}(f_s, f_d)$ corresponds to a single column of the \mathbf{A} matrix for this MTI problem.

Let us discretize the frequency variables into $N_s \geq J$ spatial frequency bins and $N_d \geq K$ Doppler frequency bins spaced uniformly across the allowed ranges for each variable to obtain $N = N_s N_d$ unique steering vectors. We can organize the resulting steering vectors into a matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$. Neglecting range ambiguities, we can define the scene reflectivity function at a given range as $\mathbf{x} \in \mathbb{C}^N$, where the rows of \mathbf{x} are indexed by angle-Doppler pairs⁸ (f_s, f_d) . We then obtain the linear relationship between the collected data and the scene of interest as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

where \mathbf{e} in this case will include the thermal noise, clutter, and other interference signals. A more realistic formulation would include measured steering vectors in the matrix \mathbf{A} . Thus, we see that the data for a multichannel pulsed radar problem can be placed easily into the framework (5.1).

5.2.2.4 Comments

The overall message is that most radar signal processing tasks can be expressed in terms of the linear model (5.1). Additional examples and detailed references can be found in [17].

We should also mention that both of these examples have used the “standard” basis for the signal of interest: voxels for SAR imaging and delay-Doppler cells for MTI. In many cases, the signal of interest might not be sparse in this domain. For example, a SAR image might be sparse in a wavelet transform or a basis of canonical scatterers.⁹ As another example, in [18], the authors explore the use of curvelets for compressing formed SAR images. Suppose that the signal is actually sparse in a basis Ψ such that $\mathbf{x} = \Psi\boldsymbol{\alpha}$, with $\boldsymbol{\alpha}$ sparse. In this case, we can simply redefine the linear problem as

$$\mathbf{y} = \mathbf{A}\Psi\boldsymbol{\alpha} + \mathbf{e} \tag{5.10}$$

to obtain a model in the same form with the forward operator $\mathbf{A}\Psi$. Indeed, many of the early CS papers were written with this framework in mind with $\mathbf{A} = \Phi\Psi$, where Φ is the measurement operator, and Ψ is the sparse basis. In this framework, one attempts to define a measurement operator Φ that is “incoherent” from the basis Ψ . We will not dwell on this interpretation of the problem and refer the interested reader to, for example, [19].

⁸Note that \mathbf{x} in this context is the same image reflectivity function used in common STAP applications [15].

⁹Selection of the appropriate sparse basis is problem dependent, and we refer the reader to the literature for more detailed explorations.

5.2.3 Regularization of the Linear Model through Sparsity

We now turn our attention to the inverse problem of estimating \mathbf{x}^{true} from the noisy data \mathbf{y} . As mentioned in Section 5.2.1, since $\delta < 1$, the problem is underdetermined. Let us consider for a moment the noise-free case with $\mathbf{e} = \mathbf{0}$. Then, $\mathbf{y} = \mathbf{A}\mathbf{x}^{true}$. Since $M < N$, the matrix \mathbf{A} has a nullspace defined as $\text{null}(\mathbf{A}) = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}$. Notice that for all \mathbf{x} in the nullspace of \mathbf{A} , we see that $\mathbf{A}(\mathbf{x} + \mathbf{x}^{true}) = \mathbf{0} + \mathbf{y} = \mathbf{y}$. Thus, we cannot hope to identify \mathbf{x}^{true} uniquely from \mathbf{y} given no other information, since there are infinitely many signals that will result in the same set of collected data.

In order to obtain a unique solution to this problem, we need to regularize the problem by adding more assumptions about the unknown signal \mathbf{x}^{true} . One common approach is to find the solution with the minimum energy or ℓ_2 norm, in which case we would estimate the signal as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \|\mathbf{x}\|_2 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{y}$$

This formulation yields a unique solution which can be computed in closed form as $\hat{\mathbf{x}} = \mathbf{A}^+\mathbf{y}$, where \mathbf{A}^+ is the Moore-Penrose pseudo-inverse¹⁰ of \mathbf{A} [20]. While a unique solution is obtained, this low energy assumption may not be reasonable for some problems of interest. To illustrate this point, consider a simple example with $N = 256$, $M = 75$, and $s = 15$. For reasons that will become clear later, the matrix \mathbf{A} contains entries drawn randomly from the Gaussian distribution $\mathcal{N}(0, 1)$. Figure 5-4(a) shows the true signal and the minimum-energy reconstruction. The ℓ_2 norm penalizes large values, so the resulting estimate contains many small coefficients. Put another way, among the infinite number of possible solutions, the ℓ_2 regularization favors solutions which use combinations of several atoms with small coefficients in favor of a few large coefficients. Unfortunately, the signal of interest has precisely this feature—a small number of large coefficients.

Our original motivation for CS techniques was to exploit signals that could be expressed with just a few nonzero coefficients. As such, it seems that we should look for the solution $\hat{\mathbf{x}}$ with the smallest ℓ_0 norm. So, we might try to solve the problem

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{y}$$

Unfortunately, the ℓ_0 norm is not differentiable or even continuous. This problem is in fact NP-hard to solve. Put another way, we would need to test every possible combination of active coefficients to verify that we have found the one with the smallest ℓ_0 norm that allows the signal \mathbf{y} to be represented. This combinatorial complexity is intractable for problems of interesting size, where \mathbf{x} may contain thousands or even millions of unknowns. As a result, we must explore practical methods for performing sparse reconstruction (SR), which is the estimation of an unknown signal that is assumed to be sparse.

¹⁰Given the SVD as $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, the Moore-Penrose pseudo-inverse is calculated as $\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^H$. Since $\mathbf{\Sigma}$ is a diagonal matrix of singular values, $\mathbf{\Sigma}^+$ is obtained by simply transposing the matrix and then taking the reciprocals of the nonzero diagonal elements. When \mathbf{A} has full column rank, the pseudo-inverse can be computed with the perhaps more familiar expression $\mathbf{A}^+ = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$.

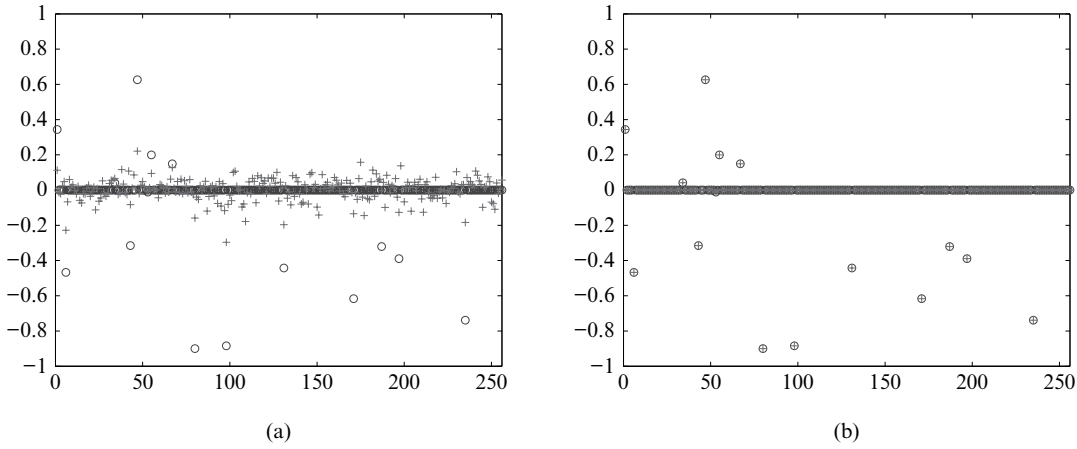


FIGURE 5-4 ■ Two reconstructions of a signal with $N = 256$, $M = 75$, and $s = 15$. The entries of A are generated randomly from a Gaussian distribution and then scaled to yield unit-norm columns. The true signal is shown with circles, while the estimate is shown with crosses. Pane (a) shows the minimum ℓ_2 reconstruction, while pane (b) shows the minimum ℓ_1 reconstruction. A very similar example was shown in [21]. The reconstructions were computed using the CVX software package [22].

5.2.4 ℓ_1 Regularization

In Section 5.3 we will explore numerous algorithms for SR. Here, we will explore a problem formulation that motivates many of these algorithms. Let us return to considering the noisy data case where $e \neq 0$ given in (5.1). In this setting, we would like to find the solution \hat{x} given by

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|x\|_0 \text{ subject to } \|Ax - y\|_2 \leq \sigma \quad (5.11)$$

However, this problem is once again NP-hard and effectively impossible to solve for problems of interest in radar signal processing. As mentioned earlier, the issue is that the ℓ_0 norm is not amenable to optimization. Figure 5-1 provides an intuitive alternative: we can replace the intractable ℓ_0 norm with a similar norm for which optimization is simpler. We have already seen that the ℓ_2 norm provides one possibility, but the resulting solutions tend to be nonsparse. Instead, we will consider the convex relaxation [23] of (5.11) using the ℓ_1 norm:

$$\hat{x}_\sigma = \underset{x}{\operatorname{argmin}} \|x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \sigma \quad (5.12)$$

We will refer to this convex optimization problem as Basis Pursuit De-Noising (BPDN).¹¹ By virtue of being a convex cost function with a convex constraint, the problem described in (5.12) does not suffer from local minima, and a variety of mature techniques exist for solving the problem in polynomial time [24]. Figure 5-4(b) shows the reconstruction of our simple example signal with an ℓ_1 penalty. In this noise-free case, the signal is reconstructed perfectly using the ℓ_1 -based cost function. Notice that this optimization problem has an obvious parameter, σ , that could be varied to obtain different solutions. We will explore this idea in depth in Section 5.3.1.1.

¹¹See Section 5.3.1.1 for details on our naming convention.

Regularization using the ℓ_1 norm has a long history, for example, [25]. We shall discuss several formulations of the problem described in (5.12) and algorithms for solving it in Section 5.3.1. When the problem is solved with an ℓ_2 penalty in place of the ℓ_1 norm, the result is termed Tikhonov regularization [26],¹² which is known in the statistics community as ridge regression [28]. This formulation has the advantage of offering a simple, closed-form solution that can be implemented robustly with an SVD [20]. Unfortunately, as in the noise-free case, this approach does not promote sparsity in the resulting solutions. We mention Tikhonov regularization because it has a well-known Bayesian interpretation using Gaussian priors. It turns out that the ℓ_1 -penalized reconstruction can also be derived using a Bayesian approach.

To cast the estimation of \mathbf{x}^{true} in a Bayesian framework, we must adopt priors on the signal and disturbance. First, we will adopt a Laplacian prior¹³ on the unknown signal \mathbf{x}^{true} and assume that the noise \mathbf{e} is circular Gaussian with known covariance Σ , that is,

$$\begin{aligned}\mathbf{e} &\sim \mathcal{CN}(0, \Sigma) \\ p(\mathbf{x}^{true}) &\propto \exp \left\{ -\frac{\lambda}{2} \|\mathbf{x}^{true}\|_1 \right\}\end{aligned}$$

where the normalization constant on $p(\mathbf{x}^{true})$ is omitted for simplicity. Given no other information, we could set $\Sigma = \mathbf{I}$, but we will keep the generality. We can then find the MAP estimate easily as

$$\begin{aligned}\hat{\mathbf{x}}_\lambda &= \operatorname{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) \\ &= \operatorname{argmax}_{\mathbf{x}} \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \\ &= \operatorname{argmax}_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \\ &= \operatorname{argmax}_{\mathbf{x}} \exp \left\{ -\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_\Sigma^2 \right\} \exp \left\{ -\frac{\lambda}{2} \|\mathbf{x}\|_1 \right\} \\ &= \operatorname{argmin}_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_\Sigma^2 + \lambda \|\mathbf{x}\|_1\end{aligned}$$

where $\|\mathbf{x}\|_\Sigma^2 = \mathbf{x}^H \Sigma^{-1} \mathbf{x}$. The resulting optimization problem is precisely what we would expect given the colored Gaussian noise prior. Since Σ is a covariance matrix, and hence positive definite and symmetric, the problem is convex and solvable with a variety of techniques. In fact, we can factor the inverse of the covariance using the Cholesky decomposition as $\Sigma^{-1} = \mathbf{R}^H \mathbf{R}$ to obtain

$$\begin{aligned}\hat{\mathbf{x}}_\lambda &= \operatorname{argmin}_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_\Sigma^2 + \lambda \|\mathbf{x}\|_1 \\ &= \operatorname{argmin}_{\mathbf{x}} \|\mathbf{R}\mathbf{A}\mathbf{x} - \mathbf{R}\mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \\ &= \operatorname{argmin}_{\mathbf{x}} \|\bar{\mathbf{A}}\mathbf{x} - \bar{\mathbf{y}}\|_2^2 + \lambda \|\mathbf{x}\|_1\end{aligned}\tag{5.13}$$

¹²An account of the early history of Tikhonov regularization, dating to 1955, is given in [27].

¹³Recent analysis has shown that, while the Laplacian prior leads to several standard reconstruction algorithms, random draws from this distribution are not compressible. Other priors leading to the same ℓ_1 penalty term but yielding compressible realizations have been investigated. See [29] for details.

where $\bar{\mathbf{A}} = \mathbf{R}\mathbf{A}$, and $\bar{\mathbf{y}} = \mathbf{R}\mathbf{y}$. This problem is equivalent to (5.12) when λ is chosen correctly, as detailed in Section 5.3.1.1. Readers familiar with adaptive processing will recognize the application of \mathbf{R} as a pre-whitening step. Indeed, this processing is the ℓ_1 version of the typical pre-whitening followed by matched filtering operation used in, for example, STAP [15,30].

Returning to the geometric interpretation of the problem, examination of Figure 5-1 provides an intuitive geometric reason that the ℓ_1 norm is effective for obtaining sparse solutions. In particular, sparse solutions contain numerous zero values and thus lie on the coordinate axes in several of their dimensions. Since the ℓ_1 unit ball is “spiky” (i.e., more pointed along the coordinate axes than the rounded ℓ_2 norm), a potential solution \mathbf{x} with zero entries will tend to have a smaller ℓ_1 norm than a non-sparse solution. We could of course consider $p < 1$ to obtain ever more “spiky” unit balls, as is considered in [31]. Using $p < 1$ allows sparse signals to be reconstructed from fewer measurements than $p = 1$, but at the expense of solving a non-convex optimization problem that could feature local minima.

This geometric intuition can be formalized using so-called tube and cone constraints as described in, for example, [1]. Using the authors’ terminology, the tube constraint follows from the inequality constraint in the optimization problem represented in equation (5.12):

$$\begin{aligned} \|\mathbf{A}(\mathbf{x}^{true} - \hat{\mathbf{x}}_\sigma)\|_2 &\leq \|\mathbf{A}\mathbf{x}^{true} - \mathbf{y}\|_2 + \|\mathbf{A}\hat{\mathbf{x}}_\sigma - \mathbf{y}\|_2 \\ &\leq 2\sigma \end{aligned}$$

The first line is an application of the triangle inequality satisfied by any norm, and the second follows from the assumed bound on \mathbf{e} and the form of (5.12). Simply put, any vector \mathbf{x} that satisfies $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \sigma$ must lie in a cylinder centered around $\mathbf{A}\mathbf{x}^{true}$. When we solve the optimization problem represented in equation (5.12), we choose the solution inside this cylinder with the smallest ℓ_1 norm.

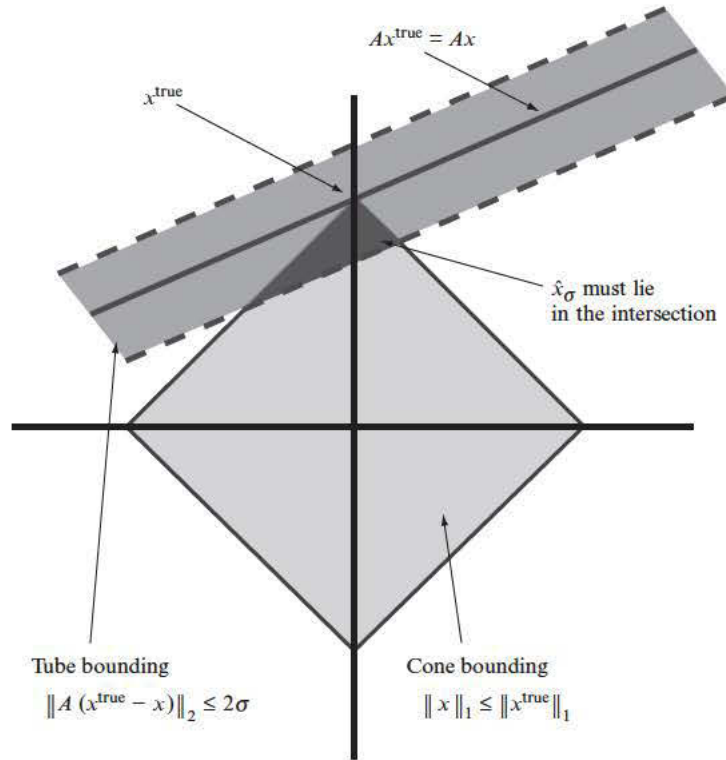
Since $\hat{\mathbf{x}}_\sigma$ is a solution to the convex problem described in (5.12) and thus a global minimum, we obtain the cone constraint¹⁴ $\|\hat{\mathbf{x}}_\sigma\|_1 \leq \|\mathbf{x}^{true}\|_1$. Thus, the solution to (5.12) must lie inside the smallest ℓ_1 ball that contains \mathbf{x}^{true} . Since this ℓ_1 ball is “spiky”, our hope is that its intersection with the cylinder defined by the tube constraint is small, yielding an accurate estimate of the sparse signal \mathbf{x}^{true} . These ideas are illustrated in two dimensions in Figure 5-5. The authors of [1] go on to prove just such a result, a performance guarantee for CS. However, sparsity of the true signal is not enough by itself to provide this guarantee. We will need to make additional assumptions on the matrix \mathbf{A} .

5.2.5 Performance Guarantees

We should emphasize at this point that SR by itself is not CS. Instead, CS involves combining SR algorithms with constraints on the measurement matrix \mathbf{A} , typically satisfied through randomized measurement strategies, to provide provable performance guarantees. In this section, we will explore several conditions on \mathbf{A} that yield performance guarantees for CS.

¹⁴The authors in [1] actually use the name *cone constraint* for a condition on the reconstruction error that is derived from this inequality.

FIGURE 5-5 ■ A two-dimensional illustration of the tube and cone constraints. This figure was adapted from Figure 5-1 in [1].



5.2.5.1 Kruskal Rank

Since we are interested in recovering sparse signals x^{true} , it would be desirable to have a condition guaranteeing that a sparse solution to our problem is unique. Let us again consider the noise-free case with $e = \underline{0}$. Our collected data then satisfy $y = Ax^{true}$, where $\|x^{true}\|_0 = s$. Suppose that there exists a second solution x_a such that $y = Ax_a$ and $\|x_a\|_0 \leq s$. We immediately see that

$$\begin{aligned} A(x^{true} - x_a) &= Ax^{true} - Ax_a \\ &= y - y = \underline{0} \end{aligned}$$

Now notice that $\|x^{true} - x_a\|_0 \leq 2s$ by assumption. We have just shown that this vector must be in the nullspace of the matrix A . Thus, if we require that $\text{null}(A)$ contains no vectors with sparsity less than or equal to $2s$, we arrive at a contradiction and conclude that the sparse solution x^{true} to $y = Ax$ must be unique. We shall formalize this notion using the idea of Kruskal rank.

We define the Kruskal rank [32] $\mathcal{K}(A)$ as the largest integer such that for all x in $\text{null}(A)$, $\|x\|_0 > \mathcal{K}(A)$. Put another way, every set of $\mathcal{K}(A)$ columns of A is full rank, but there exists at least 1 set of $\mathcal{K}(A) + 1$ columns that is linearly dependent. If $\mathcal{K}(A) \geq 2s$, then an s sparse solution to $y = Ax$ is unique. This result was provided, along with numerous additional facts, in [33].¹⁵

¹⁵Donoho and Elad [33] deal with the Spark of A , which is simply the Kruskal rank plus 1.

While the notion of Kruskal rank is important in sparse regularization, it has very limited utility for the problems of interest to radar practitioners. The regular rank of a matrix has limited utility, because an arbitrarily small change in the matrix can alter the rank. Put another way, a matrix can be “almost” rank deficient. In practice, we use measures like the condition number [20] to assess the sensitivity of matrix operations to small errors. The problem with Kruskal rank is analogous. If there exists a sparse vector such that $\mathbf{A}\mathbf{x} \approx \mathbf{0}$, this will not violate the Kruskal rank condition. However, when even a small amount of noise is added to the measurements, distinctions based on arbitrarily small differences in the product $\mathbf{A}\mathbf{x}$ will not be robust. What we need is a condition on \mathbf{A} that guarantees sparse solutions will be unique, but also provides robustness in the presence of noise. As we shall see, this condition will also guarantee successful sparse reconstruction when solving the convex relaxation of our ℓ_0 problem.

5.2.5.2 The Restricted Isometry Property

An isometry is a continuous, one-to-one invertible mapping between metric spaces that preserves distances [34]. As discussed in the previous section, we want to establish a condition on \mathbf{A} that provides robustness for sparse reconstruction. While several conditions are possible, we shall focus on the restricted isometry property (RIP). In particular, we will define the restricted isometry constant (RIC) $\mathcal{R}_n(\mathbf{A})$ as the smallest positive constant such that

$$(1 - \mathcal{R}_n(\mathbf{A})) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \mathcal{R}_n(\mathbf{A})) \|\mathbf{x}\|_2^2 \quad (5.14)$$

for all \mathbf{x} such that $\|\mathbf{x}\|_0 \leq n$. In other words, the mapping \mathbf{A} preserves the energy in sparse signals with n or fewer nonzero coefficients up to a small distortion. We refer to this condition as RIP, since the required approximate isometry is *restricted* to the set of sparse signals.

As we can see, this condition avoids the problem of arbitrarily small $\mathbf{A}\mathbf{x}$ values for sparse \mathbf{x} that can occur when only the Kruskal rank condition is required. This property is analogous to a full rank matrix having a small condition number. Indeed, the RIP can be interpreted as a requirement on the condition number of all submatrices of \mathbf{A} with n or fewer columns. Furthermore, notice that \mathbf{A} has Kruskal rank of at least n provided that $\mathcal{R}_n(\mathbf{A}) < 1$. Thus, the RIP guarantees the uniqueness of a sparse solution with any meaningful RIC. To guarantee good SR performance, a smaller RIC is required.

We are now in a position to state one of the fundamental results in CS. If $\mathcal{R}_{2s}(\mathbf{A}) < \sqrt{2} - 1$, then

$$\|\mathbf{x}^{true} - \hat{\mathbf{x}}_\sigma\|_2 \leq C_0 s^{-1/2} \|\mathbf{x}^{true} - \mathbf{x}_s^{true}\|_1 + C_1 \sigma \quad (5.15)$$

where C_0 and C_1 are small positive constants whose values and derivation can be found in [35].¹⁶ First, the term \mathbf{x}_s^{true} is the best s sparse approximation to \mathbf{x}^{true} . Thus, if \mathbf{x}^{true} is truly sparse, then the first term is zero. If the true signal is not actually sparse, then the reconstruction remains well behaved. The second term is a small multiple of the noise energy. If the measurements \mathbf{y} are noise free, then solving problem described in (5.12) with $\sigma = 0$ produces a *perfect* reconstruction of a truly sparse signal. The theorem requires a

¹⁶The reference is, in our opinion, a concise and elegant proof of this result. More detailed and perhaps pedagogically useful proofs, albeit with slightly inferior guarantees, can be found in [1]. This proof is also limited to the real case, but the extension to complex-valued signals, along with slightly less restrictive RIC results, is provided in [36].

constraint on $\mathcal{R}_{2s}(\mathbf{A})$, even though the signal of interest is assumed to be s sparse, because the proof, like the previous proof for uniqueness of the sparse solution, relies on preserving the energy of differences between s -sparse signals.

We should emphasize that this condition is sufficient but not necessary. Indeed, good reconstructions using (5.12) are often observed with a measurement matrix that does not even come close to satisfying the RIP. For example, the authors in [37] demonstrate recovery of sinusoids in noise using SR with performance approaching the Cramer-Rao lower bound for estimating sinusoids in noise, despite having the RIC for \mathbf{A} approach unity.¹⁷ Nonetheless, RIP offers powerful performance guarantees and a tool for proving results about various algorithms, as we shall see in Section 5.3. Indeed, in some cases the effort to prove RIP-based guarantees for an algorithm has led to improvements in the algorithm itself, such as the CoSaMP algorithm [39] discussed in Section 5.3.4. Other conditions exist in the literature, for example, [40–42], and indeed developing less restrictive conditions is an active area of research.

5.2.5.3 Matrices that Satisfy RIP

At this point, it may seem that we have somehow cheated. We started with an NP-hard problem, namely finding the solution with the smallest ℓ_0 norm. The result given in (5.15) states that we can instead solve the convex relaxation of this problem and, in the noise-free case, obtain the exact same solution. The missing detail is that we have assumed that \mathbf{A} has a given RIC. Unfortunately, computing the RIC for a given matrix is an NP-hard task. Indeed, it requires computing an SVD of *every* possible subset of n columns of \mathbf{A} . For a matrix of any meaningful size, this is effectively impossible.

So, it seems that we may have traded an NP-hard reconstruction task for an NP-hard measurement design task. Put another way, for our straightforward convex reconstruction problem to have desirable properties, we must somehow design a measurement matrix \mathbf{A} with a property that we cannot even verify. Fortunately, an elegant solution to this problem exists. Rather than designing \mathbf{A} , we will use randomization to generate a matrix that will satisfy our RIC requirements with very high probability.¹⁸

Numerous authors have explored random matrix constructions that yield acceptable RICs with high probability. Matrices with entries that are chosen from a uniform random distribution, a Gaussian distribution,¹⁹ a Bernoulli distribution, as well as other examples satisfy the required RIP provided that M is greater than $Cs \log(N/s)$ for some distribution-dependent constant C [21]. A very important case for radar and medical imaging applications is that a random selection of the rows of a discrete Fourier transform matrix also satisfies a similar condition with $M \geq Cs \log^4(N)$ [45]. Furthermore, \mathbf{A} can be constructed

¹⁷See also [38] for an investigation of modifying the RIP property to address performance guarantees in situations where the standard RIP is violated.

¹⁸Several attempts have been made to develop schemes for constructing \mathbf{A} matrices deterministically with the desired RIP properties, for example [47]. However, these results generally require more measurements (i.e., larger M) to guarantee the same RIC. See [43] for an example construction based on Reed-Muller codes that does not satisfy RIP for all vectors but preserves the energy of a randomly drawn sparse vector with high probability. Expander graphs have also been explored as options for constructing appropriate forward operators with accompanying fast reconstruction algorithms; see for example [44].

¹⁹Indeed, this result holds for the wider class of sub-Gaussian distributions.

as $\Phi\Psi$, where Ψ is a basis for \mathbb{C}^N , and Φ is a random matrix of one of the mentioned classes [21]. The proofs of these results rely on probabilistic arguments involving concentration of measure, see for example, [46]. Intuitively, the idea is that random vectors drawn in a very high-dimensional space are unlikely to have large inner products.

For random matrices, we have seen that we need to collect on the order of $s \log(N/s)$ measurements. A simple analog to bit counting may provide some intuition about this bound. If we have N coefficients with s nonzero values, then there are $\binom{N}{s}$ possible sets of nonzero coefficients. If we allow c_0 quantization levels for encoding each nonzero coefficient, then the total number of required bits for this information is $\log\{\binom{N}{s} + c_0 s\}$. If we neglect the bits for encoding the values, we can obtain

$$\begin{aligned} \log \left\{ \binom{N}{s} + c_0 s \right\} &\approx \log \binom{N}{s} \\ &\leq \log \left\{ \left(\frac{Ne}{s} \right)^s \right\} \end{aligned} \quad (5.16)$$

$$= s \log(N/s) + s \log e \quad (5.17)$$

Thus, a simple calculation of the required coding bits yields a leading-order term of the same form as the number of required measurements predicted by CS theory. Intuitively, randomization of the \mathbf{A} matrix ensures with high probability that each measurement provides a nearly constant increment of new information bits. This result in no way constitutes a proof but is presented to provide some intuitive insight about the origin of this expression.

We should mention that randomization has a long history in radar signal processing. For example, array element positions can be randomized to reduce the sidelobes in sparsely populated arrays [48–50]. It is also well understood that jittering or staggering the pulse repetition frequency can eliminate ambiguities [13]. The transmitted waveform itself can also be randomized, as in noise radar [51,52], to provide a thumbtack-like ambiguity function. From a CS perspective, these randomization techniques can be viewed as attempts to reduce the mutual coherence of the forward operator \mathbf{A} [17].

5.2.5.4 Mutual Coherence

As pointed out already, estimating and testing the RIC for large M is impractical. A tractable yet conservative bound on the RIC can be obtained through the mutual coherence of the columns of \mathbf{A} defined as

$$\mathcal{M}(\mathbf{A}) = \max_{i \neq j} |\mathbf{A}_i^H \mathbf{A}_j|$$

Mutual coherence can be used to guarantee stable inversion through ℓ_1 recovery [53,54], although these guarantees generally require fairly small values of s . Furthermore, the RIC is conservatively bounded by $\mathcal{M}(\mathbf{A}) \leq \mathcal{R}_s(\mathbf{A}) \leq (s-1)\mathcal{M}(\mathbf{A})$. The upper bound is very loose, as matrices can be constructed for which the RIC is nearly equal to the mutual coherence over a wide range of s values [55].

The mutual coherence is of particular importance in radar signal processing. Recall from Section 5.2.2.2 that entries of the Gramian matrix $\mathbf{A}^H \mathbf{A}$ are samples of the radar ambiguity function. The mutual coherence is simply the maximum off-diagonal of this matrix. Thus, the mutual coherence of a radar system can be reduced by designing the ambiguity

function appropriately. This view was explored for the ambiguity function, along with a deterministic approach to constructing waveforms that yield low mutual coherence for the resulting \mathbf{A} , in [56]. In a nutshell, the thumbtack ambiguity functions that are known to be desirable for radar systems [57] are also beneficial for CS applications. In [12], the authors use mutual coherence as a surrogate for RIP when designing waveforms for multistatic SAR imaging. As one might expect, noise waveforms provide good results in both scenarios.²⁰

The ambiguity function characterizes the response of a matched filter to the radar data. At the same time, the ambiguity function determines the mutual coherence of the forward operator \mathbf{A} , which provides insights into the efficacy of SR and CS. Thus, CS does not escape the limitations imposed by the ambiguity function and the associated matched filter. Note that virtually all SR algorithms include application of the matched filter \mathbf{A}^H repeatedly in their implementations. Indeed, SR algorithms leverage knowledge of the ambiguity function to approximately deconvolve it from the reconstructed signal. Put another way, SR can yield signal estimates that lack the sidelobe structure typical of a matched filtering result, but the extent to which this process will be successful is informed by the ambiguity function.

5.3 | SR ALGORITHMS

In the previous section, we surveyed much of the underlying theory of CS. As we have seen, CS combines randomized measurements with SR algorithms to obtain performance guarantees for signal reconstruction. In this section, we will review several examples of SR algorithms and their associated CS performance guarantees. It is worth mentioning that these algorithms can and are used in situations when the sufficient conditions associated with CS are not satisfied. In spite of this failure to satisfy these conditions, the resulting reconstructions are often desirable.

One issue is that the traditional CS theory provides error bounds on reconstructing \mathbf{x}^{true} . In many radar problems, the signal \mathbf{x}^{true} represents fine sampling of a parameter space, such as the set of image voxels or the angle-Doppler plane. In these scenarios, producing a reconstruction whose nonzero elements are slightly shifted in the vector $\hat{\mathbf{x}}$ may be perfectly acceptable to a practitioner, as this would correspond to a small error in estimating the relevant parameter. However, the traditional error definitions of CS would suggest that this reconstruction is extremely poor.

To give a concrete example, suppose that our signal of interest is a single tone in noise. The vector \mathbf{x}^{true} represents the discrete Fourier transform (DFT) of the signal sampled on a fine grid, and \mathbf{A} is simply a DFT matrix. If the true signal is zero except for a single entry in the first position equal to 1 and $\hat{\mathbf{x}}_\sigma$ contains a single 1 in the second position, then we have almost perfectly reconstructed the signal. The model order is correct with a single sinusoid, and the frequency has been estimated to an accuracy equal to the sampling density of our frequency grid. Yet the CS measure of error $\|\mathbf{x}^{true} - \hat{\mathbf{x}}_\sigma\|_2$ would be larger

²⁰We note that it is not possible to improve the Kruskal rank of a matrix by left-multiplication; see [58]. This motivates attempts to change the waveform or collection geometry in radar problems to improve sparse reconstruction performance rather than simply considering linear transformations of the collected data.

than the norm of the true signal, suggesting complete failure. See [37] for an excellent discussion of this issue. This example perhaps suggests why one might consider using a given SR algorithm even when conditions like RIP are not met, provided that the signal of interest is still sparse or nearly so. Performance guarantees for these sample parameter problems remain at least a partially open research problem, although some progress has been made for the specific DFT case in [59].

We shall consider several classes of SR algorithms along with examples. New algorithms are being developed and extended at a breathtaking pace in the CS literature. Indeed, it is not at all uncommon to see articles improving and extending other articles that are still available only as preprints. In some cases, multiple generations of this phenomenon are observed. Thus, while several *current* state-of-the-art algorithms will be referenced, online references can be consulted for new developments before actually employing these techniques. The good news is that a wide range of excellent SR algorithms for both general and fairly specific applications are available online for download and immediate use.²¹ Furthermore, part of the beauty of these SR algorithms is their general simplicity. Several highly accurate algorithms can be coded in just a few dozen lines in a language like MATLAB.

Before delving into the collection of algorithms, we will make a few comments about the required inputs. The SR algorithms we survey will typically require a handful of parameters along with the data y . Most of the algorithms will require either an explicit estimate of s or a regularization parameter that is implicitly related to the sparsity. In many cases, these parameters can be selected with relative ease. In addition, many of the algorithms are amenable to warm-starting procedures, where a parameter is varied slowly. The repeated solutions are greatly accelerated by using the solution for a similar parameter setting to initialize the algorithm. One of the many algorithms leveraging this idea is Nesterov's Algorithm (NESTA) [61].

Finally, an issue of particular importance is the required information about the forward operator A . Obviously, providing the matrix A itself allows any required computations to be completed. However, in many cases, A represents a transform like the DFT or discrete wavelet transform (DWT), or some other operator that can be implemented without explicit calculation and storage of the matrix A . Since the A matrix can easily be tens or even hundreds of gigabytes in some interesting problems, the ability to perform multiplications with A and A^H without explicit storage is essential. While some of the SR algorithms require explicit access to A itself, many first-order algorithms require only the ability to multiply a given vector with A and A^H . We will focus primarily on these algorithms, since they are the only realistic approaches for many large-scale radar problems.

We will divide our discussion of SR algorithms into a series of subsections. First, we will discuss penalized least squares methods for solving variants of (5.12). We will then turn to fast iterative thresholding methods and closely related reweighting techniques. All of these approaches have close ties to (5.12). In contrast, greedy methods leverage heuristics to obtain very fast algorithms with somewhat weaker performance guarantees. Finally, our discussion will briefly address Bayesian approaches to CS, methods for incorporating signal structure beyond simple sparsity, and approaches for handling uncertainty in the forward operator A .

²¹An excellent list is maintained in [60].

5.3.1 Penalized Least Squares

The convex relaxation of the ℓ_0 reconstruction problem given in (5.12) can be viewed as a penalized least squares problem. We have already seen that this problem arises in a Bayesian framework by assuming a Gaussian noise prior and a Laplacian signal prior. This approach has a long history, for example, [25], and the use of the ℓ_1 norm for a radar problem was specifically proposed at least as early as [62].

The good news is that (5.12) is a linear program for real data and a second-order cone program for complex data [1]. As a result, accurate and fairly efficient methods such as interior point algorithms exist to solve (5.12) [24]. Unfortunately, these solvers are not well suited to the extremely large \mathbf{A} matrices in many problems of interest and do not capitalize on the precise structure of (5.12). As a result, a host of specialized algorithms for solving these penalized least squares problems has been created.

This section explores several of these algorithms. It should be emphasized that solvers guaranteeing a solution to (5.12) or one of the equivalent formulations discussed herein inherit the RIP-based performance guarantee given in (5.15), provided that \mathbf{A} meets the required RIP requirement.

5.3.1.1 Equivalent Optimization Problems and the Pareto Frontier

First, we will discuss several equivalent formulations of (5.12). We will adopt the nomenclature and terminology used in [63]. In this framework, the optimization problem solved in (5.12) is referred to as basis pursuit de-noising (BPDN), or BP_σ . This problem solved in the noise-free setting with $\sigma = 0$ is called simply basis pursuit (BP), and its solution is denoted $\hat{\mathbf{x}}_{\text{BP}}$. The theory of Lagrange multipliers indicates that we can solve an unconstrained problem that will yield the same solution, provided that the Lagrange multiplier is selected correctly. We will refer to this unconstrained problem as ℓ_1 penalized quadratic program and denote it as QP_λ . Similarly, we can solve a constrained optimization problem, but with the constraint placed on the ℓ_1 norm of the unknown vector instead of the ℓ_2 norm of the reconstruction error, to obtain yet a third equivalent problem. We will use the name LASSO [64], popular in the statistics community, interchangeably with the notation LS_τ for this problem. The three equivalent optimization problems can be written as

$$(\text{BP}_\sigma) \quad \hat{\mathbf{x}}_\sigma = \underset{\mathbf{x}}{\text{argmin}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \sigma \quad (5.18)$$

$$(\text{QP}_\lambda) \quad \hat{\mathbf{x}}_\lambda = \underset{\mathbf{x}}{\text{argmin}} \lambda \|\mathbf{x}\|_1 + \|\mathbf{Ax} - \mathbf{y}\|_2^2 \quad (5.19)$$

$$(\text{LS}_\tau) \quad \hat{\mathbf{x}}_\tau = \underset{\mathbf{x}}{\text{argmin}} \|\mathbf{Ax} - \mathbf{y}\|_2 \text{ subject to } \|\mathbf{x}\|_1 \leq \tau \quad (5.20)$$

We note that a fourth problem formulation known as the Dantzig selector also appears in the literature [65] and can be expressed as

$$(\text{DS}_\zeta) \quad \hat{\mathbf{x}}_\zeta = \underset{\mathbf{x}}{\text{argmin}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{A}^H(\mathbf{Ax} - \mathbf{y})\|_\infty \leq \zeta \quad (5.21)$$

but this problem does not yield the same set of solutions as the other three. For a treatment of the relationship between DS_ζ and the other problems, see [41].

The first three problems are all different ways of arriving at the same set of solutions. To be explicit, the solution to any one of these problems is characterized by a triplet of values (σ, λ, τ) which renders $\hat{\mathbf{x}}_\sigma = \hat{\mathbf{x}}_\lambda = \hat{\mathbf{x}}_\tau$. Unfortunately, it is very difficult to map

the value of one parameter into the values for the other two. However, once a solution to one problem is available, we can calculate (to at least some accuracy) the parameters for the other two solutions.²²

First, notice that only a certain range of parameters makes sense. Consider solving BP_σ with $\sigma = \|\mathbf{y}\|_2$. The solution to this problem is obviously $\hat{\mathbf{x}}_\sigma = \mathbf{0}$. Any larger value of σ will yield the same solution. Similarly, imagine solving LS_τ with $\tau = \|\hat{\mathbf{x}}_{\text{BP}}\|_1$. (Recall that $\hat{\mathbf{x}}_{\text{BP}}$ is the solution to BP_σ with $\sigma = 0$.) In other words, this is the minimum ℓ_1 solution such that $\mathbf{A}\hat{\mathbf{x}}_{\text{BP}} = \mathbf{y}$. Any larger value of τ will produce the same solution. Thus, the solution with $\mathbf{x} = \mathbf{0}$ corresponds to a large value of λ , while the solution $\hat{\mathbf{x}}_{\text{BP}}$ corresponds to the limit of the solution to QP_λ as λ approaches zero. Values outside this range will not alter the resulting solution.

The fact that the BP solution is the limit of the solution to QP_λ is important. The algorithms that solve the unconstrained problem cannot be used to precisely compute BP solutions. Algorithms that solve QP_λ exhibit a fundamental deficiency in solving BP, as can be seen by their phase transition. See [66] for results on this issue. Notice that this problem does not arise when dealing with noisy data and solving the problem for $\sigma > 0$, as the corresponding positive λ then exists. We will emphasize recovery from noisy data throughout this chapter. In contrast, much of the CS literature centers around solving the noise-free BP problem. From a coding or compression standpoint, this makes a great deal of sense. This distinction, $\sigma > 0$ vs. $\sigma = 0$, colors our discussion, since algorithms that work beautifully for BPDN may work poorly for BP and vice versa. Indeed, an example would be the approximate message passing (AMP) algorithm [66], whose development was at least partially motivated by the inability of algorithms like Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) to solve the BP problem exactly.

We can create a plot of $\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\|_2$ versus $\|\hat{\mathbf{x}}\|_1$ which is parametrized by λ (or by τ or σ) to obtain what is known as the Pareto frontier for our problem of interest. We will denote the Pareto frontier as $\phi(\tau)$. This curve represents the minimum ℓ_2 error that can be achieved for a given ℓ_1 bound on the solution norm. Pairs above this curve are sub-optimal, and pairs below the curve are unattainable. It turns out that this curve is convex. Furthermore, for a given point on the curve, the three parameters associated with the corresponding solution $\hat{\mathbf{x}}$ are given by $\phi(\tau) = \sigma = \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\|_2$, $\tau = \|\hat{\mathbf{x}}\|_1$, and λ is related to the slope of the Pareto curve at that point [63]. In particular, the slope of the Pareto curve can be calculated explicitly from the solution $\hat{\mathbf{x}}$ at that point as

$$\phi'(\tau) = - \left\| \frac{\mathbf{A}^H \mathbf{r}}{\|\mathbf{r}\|_2} \right\|_\infty$$

where $\mathbf{r} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_\tau$ [63]. This expression is closely related to λ , which is given by $\lambda = 2 \|\mathbf{A}^H \mathbf{r}\|_\infty$, as shown in [67].²³ These results are proven and discussed in detail in [63]. Thus, much like the L-curve [68,69] that may be familiar from Tikhonov regularization, the parameter λ can be viewed as a setting which allows a tradeoff between a family of Pareto optimal solutions. An example Pareto frontier plot is shown in Figure 5-6. In the figure, we have labeled the values of the end points already discussed.

²²A good discussion of the numerical issues in moving between the parameters is provided in [61]. In a nutshell, determining λ from the solution to one of the constrained problems is fairly difficult. The other mappings are somewhat more reliable.

²³Note that the factor of 2 stems from the choice to not include a 1/2 in the definition of $\hat{\mathbf{x}}_\lambda$ in (5.19).

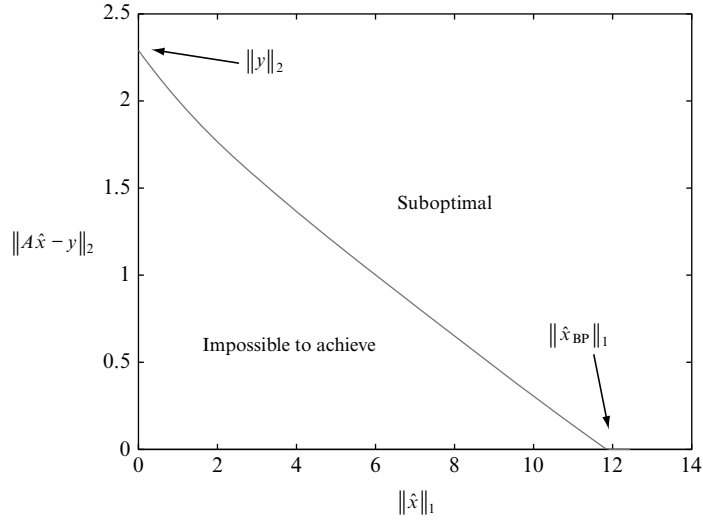


FIGURE 5-6 ■ An example of the Pareto frontier for the linear model. Points above the curve are suboptimal for any choice of the parameters, and those below the curve represent an unattainable combination of the two cost function terms. At a given point on the curve, $\sigma = \|A\hat{x} - y\|_2$, $\tau = \|\hat{x}\|_1$, and λ is related to the slope of the curve. The example was generated using the SPGL1 software [63].

If A is orthogonal, then we can approximately map $\lambda = \sigma \sqrt{2 \log N}$ [63,70]. Otherwise, it is very difficult to determine the value of one parameter given another without first solving the problem, as discussed at length in [61]. This is significant, because the parameter is often easier to choose based on physical considerations for the constrained problems, particularly BP_σ , but the constrained problems are generally harder to solve. As a result, many algorithms solve the unconstrained problem and accept the penalty of more difficult parameter selection. As already mentioned, this issue can be somewhat alleviated by solving the problem for a series of parameter values using a warm-starting or continuation approach. As we shall see in Section 5.4, the unconstrained problem is also beneficial in that we can tack on additional penalty terms to enforce various solution properties and still obtain relatively simple algorithms. Indeed, this addition of multiple penalty terms in a somewhat ad hoc, albeit effective, manner is common in practice [71–73].

Nonetheless, understanding the Pareto frontier and the relationships between the various forms of the optimization problem is highly instructive in interpreting the results. In addition, this explicit mapping between the three problems forms the foundation of the first algorithm discussed in the following section.

5.3.1.2 Solvers

In the last several years, a plethora of solvers has been created for attacking the three ℓ_1 minimization problems defined in the previous section. We will mention a handful of those approaches here. Our emphasis will be on fast algorithms that do not require explicit access to A and can handle complex-valued signals.

Our first example is SPGL1 [63], the algorithm whose primary reference inspired the discussion of the Pareto frontier in the previous section. This algorithm seeks solutions to BP_σ , which is as we have already mentioned more difficult than solving QP_λ . The algorithm

takes special advantage of the structure of the Pareto frontier to obtain the desired solution. In particular, van den Berg and Friedlander develop a fast projected gradient technique for obtaining approximate solutions to LS_τ . The goal is to approximately solve a sequence of these LASSO problems so that $\tau_0, \tau_1, \dots, \tau_k$ approaches τ_σ , which is the value for τ which renders the problem equivalent to BP_σ . While slower than solving the unconstrained problem, the fast approximate solutions to these intermediate problems allow the algorithm to solve the BP_σ in a reasonable amount of time.

Let us consider a step of the algorithm starting with τ_k . First, we compute the corresponding solution $\hat{\mathbf{x}}_\tau^k$. As discussed already, this provides both the value and an estimate of the slope of the Pareto curve as

$$\begin{aligned}\phi(\tau_k) &= \|\mathbf{A}\hat{\mathbf{x}}_\tau^k - \mathbf{y}\|_2 \\ \phi'(\tau_k) &= -\left\| \frac{\mathbf{A}^H \mathbf{r}^k}{\|\mathbf{r}^k\|_2} \right\|_\infty \\ \mathbf{r}^k &= \mathbf{A}\hat{\mathbf{x}}_\tau^k - \mathbf{y}\end{aligned}\tag{5.22}$$

We will choose the next parameter value as $\tau_{k+1} = \tau_k + \Delta\tau_k$. To compute $\Delta\tau_k$, the authors of [63] apply Newton's method. We can linearize the Pareto curve at τ_k to obtain

$$\phi(\tau) \approx \phi(\tau_k) + \phi'(\tau_k)\Delta\tau_k\tag{5.23}$$

We set this expression equal to σ and solve for the desired step to obtain

$$\Delta\tau_k = \frac{\sigma - \phi(\tau_k)}{\phi'(\tau_k)}\tag{5.24}$$

The authors of [63] provide an explicit expression for the duality gap, which provides a bound on the current iteration error, and prove several results on guaranteed convergence despite the approximate solution of the sub-problems. Further details can be found in [63], and a MATLAB implementation is readily available online. We should also mention that the SPGL1 algorithm can be used for solving more general problems, including weighted norms, sums of norms, the nuclear norm for matrix-valued unknowns, and other cases [74].

We will now discuss two closely related algorithms that were developed in the radar community for SAR imaging for solving generalizations of QP_λ . The algorithms can be used to solve the ℓ_1 problem specifically, and hence inherit our RIP-based performance guarantees, but they can also solve more general problems of potential interest to radar practitioners. First, we will consider the algorithm developed in [75] which addresses the modified cost function

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \lambda_1 \|\mathbf{x}\|_p^p + \lambda_2 \|\mathbf{D}|\mathbf{x}|\|_p^p + \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2\tag{5.25}$$

where \mathbf{D} is an approximation of the 2-D gradient of the magnitude image whose voxel values are encoded in $|\mathbf{x}|$.

This second term, for $p = 1$, is the total variation norm of the magnitude image. The TV norm is the ℓ_1 norm of the gradient. In essence, this norm penalizes rapid variation and tends to produce smooth images. As Cetin and Karl point out, this term can help to eliminate speckle and promote sharp edges in SAR imagery. Indeed, TV minimization has seen broad application in the radar, CS, and image processing communities [76]. Notice

in (5.25) that the TV norm of the magnitude of the image rather than the complex-valued reflectivity, is penalized. This choice is made to allow rapid phase variations.²⁴ Notice also that the ℓ_p norm is used with $0 < p \leq 2$. As we have mentioned, selecting $p < 1$ can improve performance but yields a nonconvex problem. Cetin and Karl replace the ℓ_p terms in the cost function with differentiable approximations, derive an approximation for the Hessian of the cost function, and implement a quasi-Newton method.

Kragh [77] developed a closely related algorithm (for the case with $\lambda_2 = 0$) along with additional convergence guarantees leveraging ideas from majorization minimization (MM).²⁵ For the case with no TV penalty, both algorithms²⁶ end up with an iteration of the form

$$\hat{\mathbf{x}}_{k+1} = [\mathbf{A}^H \mathbf{A} + h(\hat{\mathbf{x}}_k)]^{-1} \mathbf{A}^H \mathbf{y} \quad (5.26)$$

where $h(\cdot)$ is a function based on the norm choice p . The matrix inverse can be implemented with preconditioned conjugate gradients to obtain a fast algorithm. Notice that $\mathbf{A}^H \mathbf{A}$ often represents a convolution that can be calculated using fast Fourier transforms (FFTs). A more detailed discussion of these algorithms and references to various extensions to radar problems of interest, including nonisotropic scattering, can be found in [17].

These algorithms do not begin to cover the plethora of existing solvers. Nonetheless, these examples have proven useful in radar applications. The next section will consider thresholding algorithms for SR. As we shall see, these algorithms trade generality for faster computation while still providing solutions to QP_λ .

5.3.2 Thresholding Algorithms

The algorithms presented at the end of Section 5.3.1.2 both require inversion of a potentially massive matrix at each iteration. While fast methods for computing this inverse exist, we are now going to consider a class of algorithms that avoids this necessity. These algorithms will be variations on the iteration

$$\hat{\mathbf{x}}^{k+1} = \eta \{ \hat{\mathbf{x}}^k - \mu \mathbf{A}^H (\mathbf{A} \hat{\mathbf{x}}^k - \mathbf{y}) \} \quad (5.27)$$

where $\eta\{\cdot\}$ is a thresholding function. The motivation for considering an algorithm of this form becomes apparent if we examine the cost function for QP_λ . First, for ease of discussion let us define two functions that represent the two terms in this cost function. Adopting the notation used in [79]

$$\begin{aligned} f(\mathbf{x}) &= \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \\ g(\mathbf{x}) &= \lambda \|\mathbf{x}\|_1 \end{aligned}$$

²⁴Imagine, for example, the phase response of a flat plate to see intuitively why penalizing variation in the phase would be problematic.

²⁵MM relies on replacing the cost function of interest with a surrogate function that is strictly greater (hence majorization) but easier to minimize (hence minimization). The idea is to majorize the function near the current estimate, minimize the resulting approximation, and repeat. The perhaps more familiar expectation maximization algorithm is actually a special case of MM, as detailed in an excellent tutorial [78].

²⁶A particular step size must be selected to obtain this form for the algorithm in [75]. It is also worth emphasizing that this version can handle the more general problem with $\lambda_2 \neq 0$. The resulting algorithm simply uses a more complicated expression for the function h .

The function $f(\cdot)$ is differentiable, and we can easily obtain its gradient as

$$\nabla f(\mathbf{x}) = 2\mathbf{A}^H (\mathbf{A}\mathbf{x} - \mathbf{y})$$

Notice that this term appears in our generic thresholding iteration (5.27). Indeed, we can rewrite this iteration as

$$\hat{\mathbf{x}}^{k+1} = \eta \{ \hat{\mathbf{x}}^k - \mu \nabla f(\hat{\mathbf{x}}^k) \} \quad (5.28)$$

Now, the nature of the thresholding algorithms becomes obvious. At each iteration, we take a step of size μ^{27} in the negative gradient direction of the ℓ_2 portion of our cost function. We then apply a thresholding operation that tweaks these steps to account for the ℓ_1 portion of the cost function encoded in $g(\cdot)$.

5.3.2.1 Soft Thresholding

We will consider two choices for the thresholding function that yield different performance characteristics and guarantees. The first is the soft thresholding operation defined for a scalar as

$$\eta_s(x, \alpha) = \begin{cases} \left(\frac{|x| - \alpha}{|x|} \right) x, & \text{if } |x| \geq \alpha \\ 0, & \text{otherwise} \end{cases} \quad (5.29)$$

In other words, the amplitude of the scalar is reduced by α or set to zero if the amplitude is already α or less. Notice that we can apply this operation to real or complex data. When acting on a vector, the function $\eta_s(\cdot, \alpha)$ operates component-wise. The first soft threshold algorithm we will consider is the iterative shrinkage-thresholding algorithm (ISTA), which has been developed and studied by several authors. We will follow the treatment provided in [79] which demonstrates that ISTA is a MM algorithm. To see this, consider the majorization²⁸ of $f(\mathbf{x}) + g(\mathbf{x})$ at the point \mathbf{b} given by

$$Q_P(\mathbf{x}, \mathbf{b}) = f(\mathbf{b}) + \text{Re}\{\langle \mathbf{x} - \mathbf{b}, \nabla f(\mathbf{b}) \rangle\} + \frac{P}{2} \|\mathbf{x} - \mathbf{b}\|^2 + g(\mathbf{x}) \quad (5.30)$$

where P is twice the maximum eigenvalue of $\mathbf{A}^H \mathbf{A}$. This value is the smallest Lipschitz constant of $f(\mathbf{x})$, which can be easily determined using the power iteration [20] without explicit access to this potentially enormous matrix.

We can use this function to majorize our cost function at $\mathbf{b} = \hat{\mathbf{x}}^k$. The next step in defining an MM algorithm is to minimize this majorization. We can compute the unique minimizer of this function for a fixed \mathbf{b} as

$$\begin{aligned} \underset{\mathbf{x}}{\text{argmin}} \quad Q_P(\mathbf{x}, \mathbf{b}) &= \underset{\mathbf{x}}{\text{argmin}} \quad \text{Re}\{\langle \mathbf{x} - \mathbf{b}, \nabla f(\mathbf{b}) \rangle\} + \frac{P}{2} \|\mathbf{x} - \mathbf{b}\|^2 + g(\mathbf{x}) \\ &= \underset{\mathbf{x}}{\text{argmin}} \quad g(\mathbf{x}) + \frac{P}{2} \langle \mathbf{x} - \mathbf{b}, \mathbf{x} - \mathbf{b} \rangle + \frac{P}{2} \left(\left\langle \mathbf{x} - \mathbf{b}, \frac{\nabla f(\mathbf{b})}{P} \right\rangle + \left\langle \frac{\nabla f(\mathbf{b})}{P}, \mathbf{x} - \mathbf{b} \right\rangle \right) \\ &\quad + \frac{P}{2} \left\langle \frac{\nabla f(\mathbf{b})}{P}, \frac{\nabla f(\mathbf{b})}{P} \right\rangle \end{aligned}$$

²⁷The step size μ can be chosen using a variety of adaptive methods.

²⁸It is verified in [79, Lemma 2.1] that $Q_P(\mathbf{x}, \mathbf{b})$ majorizes $f(\mathbf{x}) + g(\mathbf{x})$ at \mathbf{b} .

$$\begin{aligned}
&= \underset{\mathbf{x}}{\operatorname{argmin}} \quad g(\mathbf{x}) + \frac{P}{2} \left\| \mathbf{x} - \left(\mathbf{b} - \frac{1}{P} \nabla f(\mathbf{b}) \right) \right\|_2^2 \\
&= \underset{\mathbf{x}}{\operatorname{argmin}} \quad \lambda \|\mathbf{x}\|_1 + \frac{P}{2} \left\| \mathbf{x} - \left(\mathbf{b} - \frac{2}{P} \mathbf{A}^H (\mathbf{A}\mathbf{b} - \mathbf{y}) \right) \right\|_2^2 \tag{5.31}
\end{aligned}$$

$$= \eta_s \left(\mathbf{b} - \frac{2}{P} \mathbf{A}^H (\mathbf{A}\mathbf{b} - \mathbf{y}), \frac{\lambda}{P} \right) \tag{5.32}$$

Notice that the term in parenthesis in (5.31) is a constant for fixed \mathbf{b} . Thus, the minimization in (5.31) can be carried out component-wise, yielding a simple analytical solution that corresponds to the application of the soft threshold. Combining these ideas, we obtain ISTA

$$\hat{\mathbf{x}}^{k+1} = \eta_s \left(\hat{\mathbf{x}}^k - \frac{2}{P} \mathbf{A}^H (\mathbf{A}\hat{\mathbf{x}}^k - \mathbf{y}), \frac{\lambda}{P} \right)$$

Unfortunately, ISTA has been shown to enjoy only a sublinear, that is proportional to $1/k$, rate of convergence [79, Theorem 3.1]. Beck and Teboulle [79] propose a modified fast ISTA (FISTA) that uses a very simple modification to obtain a quadratic convergence rate. FISTA requires nearly identical computational cost, particularly for a large-scale problem, and is given by

$$\begin{aligned}
\mathbf{z}^1 &= \hat{\mathbf{x}}^0 = \mathbf{0} \\
t^1 &= 1 \\
\hat{\mathbf{x}}^k &= \eta_s \left(\mathbf{z}^k - \frac{2}{P} \mathbf{A}^H (\mathbf{A}\mathbf{z}^k - \mathbf{y}), \frac{\lambda}{P} \right) \\
t^{k+1} &= \frac{1 + \sqrt{1 + 4(t^k)^2}}{2} \\
\mathbf{z}^{k+1} &= \hat{\mathbf{x}}^k + \frac{t^k - 1}{t^{k+1}} (\hat{\mathbf{x}}^k - \hat{\mathbf{x}}^{k-1})
\end{aligned}$$

Intuitively, this algorithm uses knowledge of the previous two iterates to take faster steps toward the global minimum. No additional applications of \mathbf{A} and \mathbf{A}^H are required compared with ISTA, and thus the computational cost of this modification is negligible for the large-scale problems of interest.

ISTA and FISTA²⁹ converge to true solutions of QP_λ at sublinear and quadratic convergence rates, respectively. Thus, these algorithms inherit the RIP-based performance guarantees already proven for minimization of this cost function, for example (5.15). Indeed, these algorithms are close cousins of the Kragh algorithm described in the previous section. The key difference is that the derivation is restricted to the $p = 1$ norm case to take advantage of a simple analytical result for the minimization step of the algorithm.

As a final note on these algorithms, in [61] the authors leverage the same previous work that inspired FISTA to derive the NESTA algorithm. This approach provides extremely fast computation, particularly when the forward operator \mathbf{A} enjoys certain properties. In addition, the provided algorithm can solve more general problems, including minimization

²⁹We should mention that the FISTA algorithm given in [79] is more general than the result provided here, which has been specialized for our problem of interest.

of the TV norm and nondiagonal weighting on the ℓ_1 portion of the cost function considered here. Another strong contribution of [61] is an extensive numerical comparison of several leading algorithms, including examples not considered in this chapter, on a series of test problems.

5.3.2.2 Hard Thresholding

We now turn our attention toward hard thresholding algorithms. Iterative Hard Thresholding (IHT) applies the operator $\eta_h\{\mathbf{x}, s\}$ after each gradient step. This hard thresholding function leaves the s coefficients of \mathbf{x} with the largest magnitudes³⁰ unchanged and sets all others to zero. In particular, the IHT algorithm [80] is given by

$$\hat{\mathbf{x}}^{k+1} = \eta_h\{\hat{\mathbf{x}}^k - \mu \nabla f(\hat{\mathbf{x}}^k), s\} \quad (5.33)$$

By construction *every* iteration produces a solution such that $\|\hat{\mathbf{x}}^k\|_0 \leq s$. Thus, if the algorithm parameter is set too low, we are guaranteed a priori to never find the correct solution. Naturally, this choice is analogous to the choice of λ , τ , or σ when using the ℓ_p norm-based algorithms.

A RIP-based performance guarantee for IHT is provided in [80, Theorem 4]. After we select the sparsity parameter s for IHT, then provided that $\mathcal{R}_{3s}(\mathbf{A}) < 1/\sqrt{32} \approx 0.1768$, we obtain the error bound

$$\|\mathbf{x}^{true} - \hat{\mathbf{x}}\|_2 \leq 7 [\|\mathbf{x}^{true} - \mathbf{x}_s^{true}\|_2 + s^{-1/2} \|\mathbf{x}^{true} - \mathbf{x}_s^{true}\|_1 + \sigma] \quad (5.34)$$

This guarantee is very similar to the result for BPDN. However, the RIP requirement is more stringent, requiring a condition on signals of length $3s$ instead of $2s$, and the resulting error bound is not as tight. While we do obtain a RIP condition of a similar form, it is important to note that these RIP bounds are sufficient but not necessary conditions. In addition, they are worst-case results, which may belie the performance observed for typical signals. The authors make this point at some length in [80] and provide simulations demonstrating that IHT exhibits inferior performance to BPDN and other ℓ_1 approaches when the RIP condition is violated. Thus, this simplified algorithm does come at some cost. Nonetheless, for sufficiently sparse signals, IHT performs beautifully and with excellent computational efficiency. Indeed, the need to store only a small number of nonzero coefficients at each iteration is particularly convenient for very large-scale problems. Recent work [81] has developed variations of IHT leveraging ideas from soft-thresholding schemes like FISTA with very promising numerical performance and some analytical performance guarantees.

5.3.3 Iterative Reweighting Schemes

Another class of reconstruction algorithms are iterative reweighting schemes. These techniques have a long history, for example [82,83], and have seen recent application to radar problems, for example [12,84]. The basic idea is to solve a sequence of optimization problems with a weighting matrix applied to the unknown vector. Typically, this weighting matrix is obtained from the previous iteration. These algorithms can typically be derived using an MM framework, although this perspective is perhaps less intuitive, and in some cases end up with iterative schemes that are very similar to the Kragh and Cetin

³⁰Equal-amplitude ties can be broken lexicographically.

approaches discussed in Section 5.3.1.2. An analysis and discussion of one of these iterative reweighted least squares (IRLS) approaches in terms of RIP is provided in [85].

One class of IRLS algorithms solves a sequence of ℓ_2 -regularized problems with a weighting matrix derived from the previous iterate. An example is the focal underdetermined system solver (FOCUSS) algorithm [86], which predates the work on CS. Here, we provide an example of an algorithm that solves a sequence of reweighted ℓ_1 problems as proposed in [87]. In particular, the noisy-data version of the algorithm is given as

$$\hat{\mathbf{x}}^{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{W}^k \mathbf{x}\|_1 \text{ subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \sigma$$

$$\mathbf{W}_{i,i}^k = \frac{1}{|\hat{\mathbf{x}}_i^k| + \varepsilon}$$

In other words, \mathbf{W}^k is a diagonal matrix with elements equal to the inverses of the amplitudes of the elements of the previous iterate³¹. The next estimate of \mathbf{x} is obtained by solving a new BP_σ problem with this weighting matrix applied to \mathbf{x} . As a coefficient becomes small, the weight applied to it will become very large, driving the coefficient toward zero. In this way, the reweighting scheme promotes sparse solutions. Indeed, this approach can reconstruct sparse solutions with fewer measurements than a straightforward BP_σ approach. Unfortunately, the price for this performance improvement is the need to solve a sequence of ℓ_1 optimization problems. These subproblems can be solved with one of the techniques we have already discussed.

This algorithm can be derived from a MM framework where the cost function to be minimized is the sum of the logarithms of the absolute values of the coefficients of \mathbf{x} [87]. This function's unit ball is more similar to the ℓ_0 norm than that of the ℓ_1 norm, intuitively explaining the improved performance. This approach can be extended to nonconvex minimization with $p < 1$; see [88].

5.3.4 Greedy Methods

We now turn our attention to a somewhat different class of algorithms for sparse reconstruction. Greedy algorithms are principally motivated by computational efficiency, although they often exhibit somewhat inferior performance to the approaches already considered. They do not arise from optimizing an ℓ_p -penalized cost function but instead rely on iterative attempts to identify the support of the unknown vector \mathbf{x}^{true} . Notice that if we know a priori which elements of \mathbf{x}^{true} are nonzero, indexed by the set Γ , then we can solve the much easier problem

$$\mathbf{y} = \mathbf{A}_\Gamma \mathbf{x}_\Gamma^{\text{true}} + \mathbf{e} \quad (5.35)$$

where subscripting by Γ indicates that we have thrown away the entries or columns not included in Γ . Since \mathbf{x}^{true} is sparse, the matrix \mathbf{A}_Γ now has many more rows than columns. Put simply, this problem is overdetermined and can be solved easily using least squares. In particular, we can estimate \mathbf{x}^{true} as³²

$$\hat{\mathbf{x}}_\Gamma = (\mathbf{A}_\Gamma^H \mathbf{A}_\Gamma)^{-1} \mathbf{A}_\Gamma^H \mathbf{y} \quad (5.36)$$

³¹The ε factor is included to ensure finite weights.

³²Naturally we set entries outside the set Γ to zero.

One can argue that this so-called *oracle* solution is the best that we can hope to do given knowledge of Γ and no other information, see, for example [1]. Given additional knowledge about \mathbf{e} or \mathbf{x}^{true} , we might be able to improve this estimate, but the basic idea remains that knowledge of the support set greatly simplifies our problem.

Greedy algorithms attempt to capitalize on this notion by identifying the support set of \mathbf{x}^{true} in an iterative manner. These techniques are far older than CS, tracing back to at least an iterative deconvolution algorithm known as CLEAN [89], which is equivalent to the more-recent Matching Pursuits (MP) algorithm [90]. Here, we will describe the Orthogonal Matching Pursuits (OMP) algorithm [91] as a typical example of greedy methods. As already suggested, OMP computes a series of estimates of the support set denoted by Γ_k .

OMP is based on the idea of viewing \mathbf{A} as a dictionary whose columns represent potential basis elements that can be used to represent the measured vector \mathbf{y} . At each iteration, the residual error term $\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^k$ is backprojected or multiplied by \mathbf{A}^H to obtain the signal $\mathbf{q}^k = \mathbf{A}^H(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^k)$. The largest peak in \mathbf{q}^k is identified and the corresponding atom is added to the index set, that is $\Gamma_{k+1} = \Gamma_k \cup \text{argmax}_i |\mathbf{q}_i^k|$. The new estimate of the signal is then computed as $\hat{\mathbf{x}}^{k+1} = (\mathbf{A}_{\Gamma_{k+1}}^H \mathbf{A}_{\Gamma_{k+1}})^{-1} \mathbf{A}_{\Gamma_{k+1}}^H \mathbf{y}$.

Basically, at each iteration OMP adds to the dictionary the atom that can explain the largest fraction of the energy still unaccounted for in the reconstruction of $\hat{\mathbf{x}}$. Figure 5-7 provides an example that may clarify this process. The example involves a single radar pulse with 500 MHz of bandwidth used to reconstruct a range profile containing three point targets. The initial set Γ_0 is empty. The top left pane shows the plot of $\mathbf{A}^H \mathbf{y}$. The peak of this signal is selected as the first estimate of the signal \mathbf{x}^{true} , and the corresponding amplitude is computed. The resulting estimate is shown on the top right of the figure. Notice that the estimate of the amplitude is slightly off from the true value shown with a cross. At the next iteration, the dominant central peak is eliminated from the backprojected error, and the algorithm correctly identifies the leftmost target. Notice in the middle right pane that the amplitude estimate for the middle target is now correct. OMP reestimates all of the amplitudes at each iteration, increasing the computational cost but improving the results compared with MP. The final iteration identifies the third target correctly. Notice that the third target was smaller than some of the spurious peaks in the top left range profile, but OMP still exactly reconstructs it after only three iterations.

For very sparse signals with desirable \mathbf{A} matrices, OMP performs well and requires less computation than many other methods. Unfortunately, as discussed in detail in [91,92], the performance guarantees for this algorithm are not as general as the RIP-based guarantees for algorithms like IHT or BPDN. Indeed, the guarantees lack uniformity and only hold in general in a probabilistic setting. Two issues with the OMP algorithm are perhaps responsible for these limitations. The first is that the algorithm only selects a single atom at each iteration. The second, and perhaps more crucial limitation, is that the algorithm cannot correct mistakes in the support selection. Once an atom is in the set Γ , it stays there.

Two virtually identical algorithms—compressive sampling matching pursuits (CoSaMP) [39] and subspace pursuit (SP) [93]—were developed to overcome these limitations. They select atoms in a large group at each iteration and include a mechanism for removing atoms from the dictionary when they are found to be redundant.³³ We can then

³³These algorithms are closely related to hard-thresholding approaches. Indeed, many of the SR algorithms being developed blur the lines between the classes of algorithms we have used throughout our discussion.

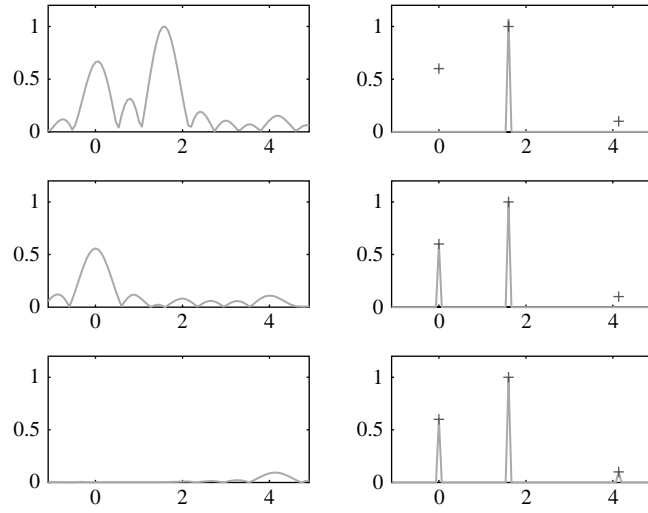


FIGURE 5-7 ■ An example application of OMP to a simple radar problem. The signal of interest is a set of three-point scatterers. The collected data represent the returns from a single pulse with 500 MHz of bandwidth. Each row depicts a single iteration of the OMP algorithm. The panes on the left depict the backprojected residual $A^H(y - A\hat{x})$ at each iteration. The right panes depict the signal estimate \hat{x} with a curve and the true signal \mathbf{x}^{true} with crosses.

obtain a RIP-based guarantee of the same form as others that we have seen. In particular, if $\mathcal{R}_{4s}(\mathbf{A}) \leq 0.1$ [39], then

$$\|\mathbf{x}^{true} - \hat{\mathbf{x}}\|_2 \leq 20 [\|\mathbf{x}^{true} - \mathbf{x}_s^{true}\|_2 + s^{-1/2} \|\mathbf{x}^{true} - \mathbf{x}_s^{true}\|_1 + \sigma] \quad (5.37)$$

In addition, CoSaMP will converge, assuming exact arithmetic, in at most $6(s + 1)$ iterations. This guaranteed convergence and numerical simplicity come at a price. In particular the requirement on the RIC is more stringent and the error bound is looser. A comparison of IHT, CoSaMP, and BP for noise-free signals can be found in [94]. A detailed numerical investigation along with parameter tuning strategies comparing CoSaMP and SP to iterative thresholding algorithms can be found in [95]. As we shall see one advantage of these greedy approaches is that they can be easily extended to cases where signal knowledge beyond simple sparsity is available.

5.3.5 Bayesian Approaches

We have already mentioned the interpretation of \mathbf{QP}_λ as the MAP estimate under a Laplacian prior. Several efforts have been made to apply other priors to the vector \mathbf{x}^{true} to achieve sparse solutions. Many of these approaches seek minimum mean square error (MMSE) estimates of the signal.

5.3.5.1 Averaging Solutions

The first example along these lines is a randomized version of OMP described in [96]. Instead of always selecting the atom having the largest correlation with the remaining residual, this algorithm selects the next atom at random with a distribution whose

probability mass values are proportional to the magnitudes of the backprojected residual. The algorithm is repeated several times, yielding a different solution at each run, and the solutions are then averaged. This approach is shown to produce a better mean square error (MSE) than straightforward OMP.

A little reflection reveals that this approach is attempting to approximate the MMSE estimate. The true MMSE estimate would be a weighted sum of all possible reconstructions scaled by their posterior probabilities. This algorithm avoids the calculation of these posterior probabilities by seeking only a few high probability solutions and assuming that they are of roughly equal likelihood. In [97], the authors pursue a somewhat more refined version of this idea. In particular, they adopt a prior on \mathbf{x}^{true} which consists of a Bernoulli indicator variable that determines whether each coefficient is nonzero. When a coefficient is nonzero, its amplitude is assumed to be Gaussian. This Bernoulli-Gaussian prior allows the authors to derive an approximation to the MMSE estimate where multiple candidate solutions are obtained through a repetitive greedy tree search and then averaged together with approximations to their posterior weights. The resulting algorithm compares favorably to a variety of existing CS algorithms and generalizes the results in [96]. In addition, it can be combined with a generalized EM scheme to estimate the hyper parameters associated with the Bernoulli and Gaussian distributions.

5.3.5.2 Graphical Models

A recent trend in CS is to examine algorithms based on belief propagation [98,99]. Belief propagation involves computing solutions based on message passing in graphical models that represent the relationships between priors and measurement variables; see [100] for an excellent overview. An overview of the application of these methods to CS problems can be found in [101].

In reality, as already briefly mentioned, RIP-based performance guarantees are often fairly conservative and always sufficient, rather than necessary, conditions. Indeed, typical algorithm performance will often exceed these guarantees even when the RIP property is not satisfied. As a result, many references report algorithm performance in terms of phase transition plots derived from Monte Carlo simulations. As briefly mentioned earlier, these plots show the performance of an algorithm for large values of N in terms of δ and ρ . The origin of the name lies in the fact that for a fixed value of δ the performance of an algorithm will often exhibit a very sharp transition as ρ is increased. For example when we fix δ in the noise-free setting, BP will recover signals with nearly 100% probability up to some critical value of ρ , as shown in Figure 5-8. For larger ρ values, the reconstruction will fail with probability 1 [102]. In the noisy setting, similar plots can be made depicting isocurves of constant MSE or other quantities. See [103] for a discussion of phase transitions with noisy data.

Development of these algorithms is ongoing. We will briefly mention one recent example, approximate message passing (AMP) [66,104–106]. It turns out that ℓ_1 optimization obtains a phase transition behavior for the noise-free case that is not matched by iterative thresholding algorithms.³⁴ AMP adds a small correction term to the ISTA iteration that

³⁴Recall that ISTA and FISTA provide solutions to BPDN provided that $\sigma > 0$. Since ISTA/FISTA solve an unconstrained problem, they cannot solve the equality-constrained case reliably.

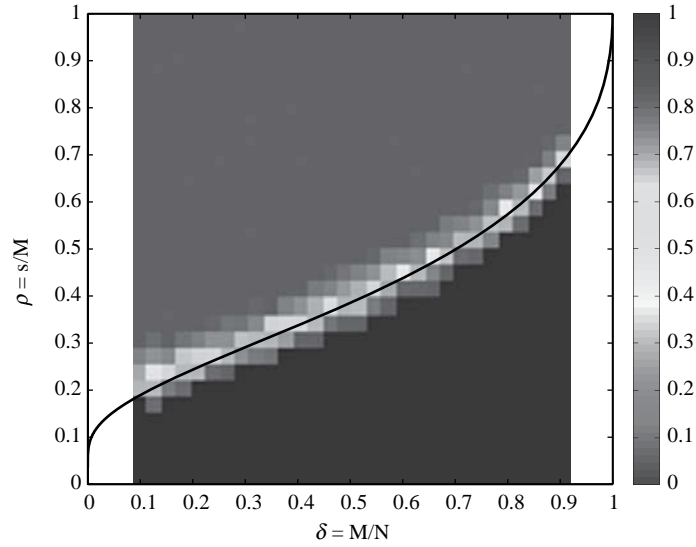


FIGURE 5-8 ■ The figure depicts the probability of exactly reconstructing an unknown signal from noise-free measurements using BP as a function of $\delta = M/N$ and $\rho = s/M$. The simulation was done for the small value $N = 256$ and only 100 Monte Carlo runs to emphasize that this behavior manifests at relatively short signal sizes. Notice the sharp transition from almost certain success to almost certain failure. The black line is the theoretical prediction. See [102] for similar plots and extensive discussion. The reconstructions were computed using the CVX software package [22].

restores the phase transition performance of BP. The algorithm is termed approximate message passing because it is derived from a belief propagation approach.

The most interesting aspect of this work is that belief propagation techniques normally rely on sparsely connected graphs to converge. AMP instead leverages the densely connected graph characteristic of a CS problem to justify approximating the messages passed on the graph as Gaussian. A series of simplifications then yields the AMP algorithm, which is nearly equivalent to ISTA except for the inclusion of an additional correction term within each iteration. The additional term has a history in the statistical physics community, as discussed in the references; see [107] for a detailed discussion. There has also been work on applying these techniques in a noisy setting, for example [103]. AMP has also been generalized to handle very general signal and measurement models with the Generalized AMP (GAMP) [108]. The graphical models approach to CS problems allows the incorporation of rich prior information including, for example, structured sparsity.

5.3.6 Structured Sparsity

In our discussions up to this point, we have focused on reconstructing signals \mathbf{x}^{true} that are sparse (compressible) in some known dictionary \mathbf{A} . One of the most promising research directions in CS is to incorporate additional information about the signal to further regularize the problem, providing additional robustness to noise and permitting accurate reconstructions with ever fewer measurements. Intuitively, one might expect that possessing additional knowledge about the permissible sparse signals might simplify signal reconstruction. An excellent analogy can be made to error-correcting codes [109], wherein

only a subset of vectors that are well separated is used for communication. This separation allows the symbols to be accurately distinguished by the receiver. CS techniques that leverage structured sparsity operate on a similar principle.

Before delving into more elaborate approaches, we mention that the simplest prior knowledge that we might exploit would be differing prior probabilities on which coefficients of \mathbf{x}^{true} are nonzero. This information could easily be available in various radar scenarios and other problems. For example, in a surveillance application looking for vehicles on a road network, it might make sense to drastically reduce the probability of target presence at large distances from roadways. This approach can be implemented by replacing λ with a diagonal weighting matrix when solving QP_λ . This modification can be incorporated into any CS algorithm; see problem 9. Notice that this scheme is closely related to the iterative reweighting reconstruction algorithms. Indeed, the authors in [86] suggest incorporating prior information into the reconstruction weights. As another example, in [110] the authors determine optimal weighting schemes for a signal with two groups of coefficients that share different prior probabilities.

However, we can also incorporate information about the *structure* of the sparsity pattern rather than simply assigning weights to specific elements. An excellent intuitive example of this idea is found in [111]. The authors seek to detect changes between images. Naturally, these change detection images will be sparse. However, they will also be clumpy, in that the nonzero coefficients will tend to occur in tightly clustered groups. The authors use a Markov random field as a prior on \mathbf{x}^{true} to promote clustering of nonzero coefficients. Motivated by the Markov model, they propose an intuitive greedy algorithm known as LaMP. The algorithm is reminiscent of CoSaMP. The primary difference is that rather than simply pruning to enforce the desired sparsity after computing the error term, LaMP uses the known structure of the sparsity pattern to determine the most likely support set for the nonzero coefficients. LaMP performs significantly better than CoSaMP on signals that satisfy this structural assumption.

Baraniuk et al. [112] extend the RIP framework to deal with structured sparse signals and provide RIP-like performance guarantees for modified versions of CoSaMP and IHT. Like the LaMP algorithm, the key to the improved performance is the exploitation of additional signal structure. For some classes of structured sparsity, the algorithm is able to reconstruct sparse signals from order s measurements, as opposed to the order $s \log(N/s)$ required by algorithms that focus on simple sparsity. In particular, their approach is developed for signals characterized by wavelet trees and block sparsity. See [113] for additional discussion of these ideas with connections to graphical models. Block sparsity is also exploited in [114]. One can also consider the joint sparsity of multiple signals as in [115]. In [116], the authors address extensions of the RIP condition to address block-sparse signals and develop recovery algorithms for them.

On a final note, [117] combines belief-propagation-based CS algorithms with ideas from turbo equalization [118] to develop an innovative framework for incorporating structured sparsity information. The algorithm alternates between two decoding blocks. The first block exploits the structure of the measurements, while the second block leverages the known structure of the sparsity patterns. By exchanging soft decisions, that is probabilities that individual elements of \mathbf{x}^{true} are nonzero, these two blocks cooperatively identify the support of the true signal. The paper demonstrates excellent results for simulations using data derived from a simple Markov chain model. This approach appears promising for incorporating structural and prior information into SR and related inference tasks.

As an aside, SR can be extended beyond the case of a sparse vector or set of jointly sparse vectors. One intriguing example is the case of reconstructing low-rank matrices from sparse sampling of their entries, for example [119], as inspired by the so-called Netflix problem.³⁵ In a related paper [120], the authors reconstruct a matrix as a sum of a low rank and an entry-wise sparse matrix. This decomposition can remove impulsive noise and was very effective for various image processing applications. In [121], a fast algorithm is proposed that can handle this decomposition task in the presence of noise and missing data. As a final example, [122] estimates the covariance matrix of a data set by assuming that the matrix's eigen-decomposition can be represented as product of a small number of Givens rotations.

5.3.7 Matrix Uncertainty and Calibration

Perfect knowledge of the forward operator \mathbf{A} cannot reasonably be expected in many radar applications, where \mathbf{A} may include assumptions about calibration, discretization, and other signal modeling issues. While additive noise can account for some of these effects, the impact of matrix uncertainty should be given specific attention. Specifically, we are interested in problems of the form

$$\mathbf{y} = (\mathbf{A} + \mathbf{E}) \mathbf{x}^{true} + \mathbf{e} \quad (5.38)$$

where the multiplicative noise \mathbf{E} is unknown. Notice that this additional error term can account for a wide range of signal modeling issues, including calibration, grid error,³⁶ autofocus, sensor placement, and manifold errors. Since the system measurement model is linear, it is perhaps unsurprising that a relatively straight forward extension to the existing RIP theory can offer limited performance guarantees in the presence of multiplicative error \mathbf{E} [126]. A simple argument in this direction can be made by bounding \mathbf{E} and absorbing its effect into the additive noise \mathbf{e} .

However, as observed in [124], basis mismatch and grid errors can lead to significant performance loss. Thus, algorithms that can compensate for matrix uncertainty directly have been developed. In ℓ_2 regularized reconstruction, the approach of total least squares is often employed to cope with matrix uncertainty. In [125], the authors extend this idea to sparse regularization with an algorithm known as sparsity-cognizant total least squares (STLS). They provide a low-complexity algorithm that can also cope with some forms of parametric matrix uncertainty. In [127], the authors propose an algorithm closely related to the DS that offers promising performance in Monte Carlo trials. Our own work in [128] leverages GAMP to address matrix uncertainty, including a parametric approach with close ties to the work in [125]. Particularly as CS is applied to a wider range of practical

³⁵Netflix created a competition for the development of algorithms for matrix completion in support of its efforts to predict customer preferences from a small sample of movie ratings.

³⁶CS uses a discretized linear model to circumvent the model order selection problems in traditional nonlinear parametric approaches, but the potential for grid error is a consequence of the underlying discretization. Specifically, in parameter estimation, the columns of \mathbf{A} represent the system response to various values of a parameter sampled on some discrete grid, for example pixels in an image or frequencies of sinusoids. When the true parameter values lie between these samples, the true signal is not perfectly represented in the dictionary; indeed, the representation may not even be sparse. See [37,123–125] for discussions of this topic.

problems, algorithms for matrix uncertainty and closely related problems in dictionary learning [129] will continue to multiply.

5.3.8 Final Thoughts

We have considered a wide range of algorithms for solving SR problems and a variety of associated CS performance guarantees. The selection of an algorithm for a given problem will be controlled by a variety of factors including at least the presence or absence of noise, the specific desired penalty terms (e.g., ℓ_1 , ℓ_p , TV), prior knowledge on the noise power and sparsity level, scale of the problem, explicit access to \mathbf{A} , availability of fast operators for \mathbf{A} and \mathbf{A}^H , desired accuracy, and the required computational speed. Each algorithm offers an operating point in this vast trade space, and it has been our goal to illustrate a few examples of the possibilities. Additional algorithms, including approaches tailored to specific problems, abound in the burgeoning CS literature. As a single example, a variety of algorithms based on the Bregman distance offer exceptional performance on a wide range of problems [130,131].

Many of the performance guarantees in the literature are only sufficient conditions. There are at least three-potential pitfalls to relying on these guarantees. First, they are often conservative, stemming from the required assumptions in the proofs and also from the fact that the analysis is often for a worst-case scenario. Typical signals may exhibit much better performance in practice. The reader is encouraged to conduct simulations, perhaps to generate a phase transition plot, to evaluate algorithm performance on a specific problem. See [42] for a discussion of attempts to compute sharp performance bounds for some CS problems. Second, as highlighted in our discussion on structured sparsity, the incorporation of additional prior and structural information can allow reconstruction performance that significantly exceeds the predictions of the RIP-style bounds. Finally, many SR tasks of interest in radar involve reconstructing a signal \mathbf{x}^{true} that represents an oversampling of a parameter space, such as a dense sampling of frequency when attempting to reconstruct sinusoids in noise. Both empirical [37] and some analytical [38] evidence suggests that SR algorithms can yield meaningful results in these scenarios despite the complete failure of these \mathbf{A} operators to satisfy the RIP and mutual coherence requirements associated with CS performance guarantees.

5.4 | SAMPLE RADAR APPLICATIONS

An abundance of radar applications can be found in which SR and CS may prove to be useful; we mention a small portion of the important work in this area and refer the reader to the Compressive Radar section of the Rice University CS website [132] for the latest research. Data link issues are common in airborne SAR applications; specifically, bandwidth constraints often require significant compression of the acquired phase history before data can be transmitted to a ground processing station for imaging and target detection. Many of the proposed techniques follow the sense-then-compress paradigm, such as Gurbuz et al. [133], who provide a scheme for random undersampling in range-frequency for ground-penetrating radar (GPR) imaging, the work of Bhattacharya et al. [134], which provides a range-frequency and Doppler-frequency undersampling scheme, and Novak's approach [135], which transforms all of the data into the biorthogonal 4.4 wavelet domain before quantizing and randomly undersampling the resulting wavelet

coefficients.³⁷ On the other hand, Patel et al. [136] compare various randomized slow-time undersampling schemes, which advantageously enable the SAR to reduce its overall duty cycle by skipping pulses.

Another potential problem is in the direct measurement of wideband signals; when the narrowband approximation is insufficient, standard baseband processing cannot be used to ease analog-to-digital conversion (ADC); thus many state-of-the-art radars are beginning to push the limits of ADC hardware. Baraniuk [137] and Romberg [138] suggest the possibility of randomly undersampling each pulse in fast time, thereby reducing the load on the ADC with the assumption that each pulse is compressible in a known basis such as the point-scatterer dictionary. For an example paper considering analog, that is, infinite dimensional signals, from a CS perspective, see [139].

A substantial percentage of radar problems are concerned with obtaining better performance (in terms of, for example, target detections, image resolution) with the same data; that is, the waveforms and acquisition strategy have already been chosen and there is no need to throw away precious data, but we still would like to exploit the underlying data sparsity for a given problem. An example of such work is the PhD research of Cetin [75], which employed various non quadratically constrained reconstructions schemes (TV, ℓ_p , etc.) in order to improve SAR image quality with data collected from DARPA's Moving and Stationary Target Acquisition and Recognition (MSTAR) program [140]. Many of the application-inspired SR techniques exploit structured sparsity. Significant examples include the work of Varshney et al. [141], which develops a greedy approach for reconstructing radar signals composed of anisotropic scattering centers, Cevher et al. [113], who employ quadtree graphical models for sparse reconstruction of SAR images, as well as Duarte and Baraniuk [59], who have developed MUSIC-based techniques that significantly outperform standard SR algorithms for the problem of estimating sinusoids in noise.

It should be noted that one often desires a measure of confidence in a given image. As discussed in Section 5.3.5, the Bayesian framework enables us to estimate posterior probabilities for reconstructed images. Algorithms for obtaining radar images with confidence labels include, for example, the fast Bayesian matching pursuits (FBMP) approach by Schniter et al. [97] and Bayesian compressive sensing (BCS) algorithm by Ji et al. [142].

Here we illustrate five different radar applications in which SR can improve image quality, the fifth of which is an application of CS. The first is a moving-target imaging example in which the target is undergoing random motion that is representative of the relative motion between a target moving down a dirt road and a SAR antenna flying along a planned linear trajectory. The second example demonstrates the use of SR in the case of a civilian vehicle; in order to efficiently model the data, dictionary elements with an anisotropic angular response are required. The third example shows how the conglomeration of physically meaningful regularization terms (ℓ_1 , ℓ_2 , and TV) can obtain excellent imaging results for imaging underground scenes. The last imaging example shows how SR can be utilized in the nonlinear setting, that is, without invoking either of the Born or Rytov approximations, by solving the far-field relation used in the linear sampling method (LSM). The final example demonstrates how the transmitted waveform can be designed to make the scattered data more amenable to sparse reconstruction in the range-Doppler plane.

³⁷It is perhaps useful to note that [135] gauged the image-compression performance by the quality of coherent change detection (CCD) imagery.

We again emphasize that radar applications of CS are becoming increasingly common in the literature. The reader is encouraged to consult the literature for additional applications. One area of particular potential interest is the use of CS to estimate time varying signals, for example [143–145]. By leveraging temporal characteristics of a signal, these methods can reduce the number of required measurements at any given time instant for reconstructing the complete time-varying signal. These ideas can be applied to radar problems in tracking and state estimation.

5.4.1 Moving Target Imaging

We will use the AFRL’s publicly released Backhoe [146] data set to illustrate the ability of SR to combat the ill-posed nature of 3-D reconstruction from sparse, narrow-angle data representative of phase history observed from ground moving targets. This data set is available upon request from the AFRL Virtual Distributed Laboratory (VDL) website [147]. The data include monostatic returns from a construction backhoe in free space (exocutter). The subset of range-compressed data we consider has a center frequency of 9.6 GHz, 640 MHz of range-frequency bandwidth, and 64 range bins per pulse. The target is assumed to be moving slow enough for us to neglect within-pulse target motion and invoke the start–stop approximation. Furthermore, we assume that the motion of the target has already been determined through an autofocus routine capable of reconstructing arbitrary target motion [148,149] to concentrate on the SR aspects of the problem. The relative target–antenna motion gives rise to the smoothly varying path of illumination directions shown in Figure 5-9 (a). The standard backprojection image $A^H y$ is shown against the target’s CAD model in Figure 5-9 (b).

The scattering phenomenology of man-made targets at microwave frequencies makes the application of SR techniques, in the context of point-scatterer reflectivity values, a sensible approach. Specifically, the relatively large size of radar targets compared with the wavelengths involved allows the data to be parsimoniously represented by a small number

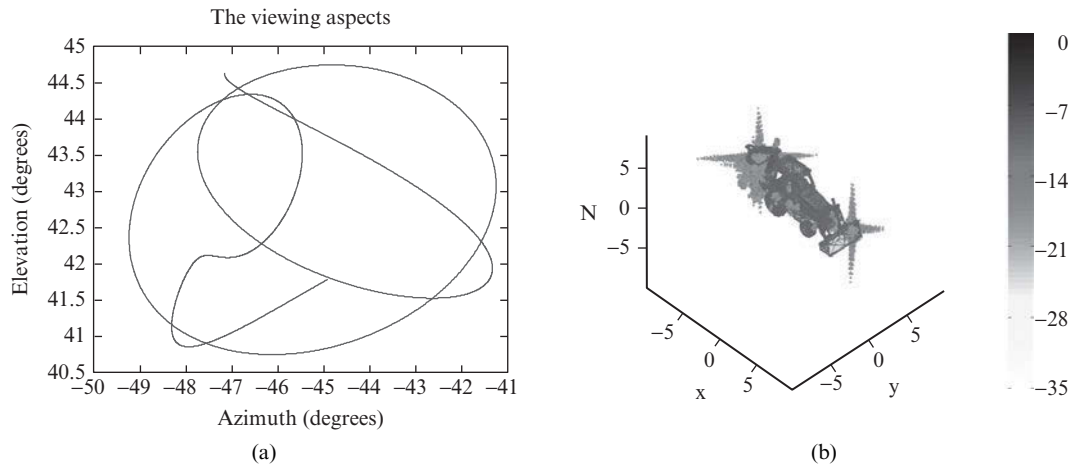


FIGURE 5-9 ■ (a) A possible nonquadratic flight path representative of one induced by a ground-moving target. (b) The dB-scaled initial image $A^H y$ formed from the limited backhoe data (view is from the nominal radar boresight). The backprojection operator A^H was calculated with a NUFFT.

of spatially distributed point-like scatterers [150]. Parametric models based on Keller's geometric theory of diffraction (GTD) have been shown to give very accurate sparse representations of radar data [151]. These GTD-based models characterize the scattered electric field E as a discrete sum of M scattering centers. In the far field, the predicted monostatic response is of the form

$$E(k) = \sum_{n=1}^N A_n (jk/k_c)^{\alpha_n} \frac{e^{2jk(\|\mathbf{t}\|_2 - \hat{\mathbf{t}} \cdot \mathbf{q}_n)}}{16\pi^2 \|\mathbf{t}\|_2^2} \quad (5.39)$$

where k_c is the center wavenumber, $\hat{\mathbf{t}} = \mathbf{t}/\|\mathbf{t}\|_2$ is the direction of the plane wave, $\mathbf{q}_n \in \mathbb{R}^3$, and each $\alpha_n \in \mathbb{Q}$ is a shape-specific parameter for a canonical target. Formulas for the complex amplitude A_n of several different canonical targets in the bistatic setting are given in [152,153], and examples of sparse regularization with these more-general GTD-based parametric models can be found in [141,154]. In the following example we use point scatterers to populate the matrix \mathbf{A} , which corresponds to setting the geometry-specific parameter $\alpha = 0$ in (5.39). Figure 5-10 compares the backprojection image (left) to the image obtained by using the FISTA algorithm to solve QP_λ with an $M \times N$ dictionary \mathbf{A} of equispaced point scatterers (right). The image is $N = 128^3 = 2097152$ pixels, and the total number of measurements is $M = (\text{number of frequencies per pulse})(\text{number of pulses}) = (64)(65\,625) = 3.6 \times 10^5$. The dictionary \mathbf{A} has a mutual coherence of 0.8142.

We mention two convenient facts regarding efficient computation of the operations $\mathbf{A}^H \mathbf{y}$, $\mathbf{A} \mathbf{x}$, and $\mathbf{A}^H \mathbf{A} \mathbf{x}$. First, in the case of far-field point scatterers, the columns of \mathbf{A} are of the form

$$\mathbf{A}_{m,n} \propto e^{i\xi_m \cdot \mathbf{q}_n} \quad (5.40)$$

where we have used the notation $\xi_m = 2k_m \hat{\mathbf{t}}_m$ and have neglected the constant amplitude terms in the denominator of (5.39). From (5.40) it is clear that \mathbf{A} is an irregularly sampled Fourier matrix because although the image locations \mathbf{q}_n may sometimes be chosen to be equispaced, the “k-space” data locations ξ_m in general are nonuniform. Thus, the forward- and backward-projection operations $\mathbf{A} \mathbf{x}$ and $\mathbf{A}^H \mathbf{y}$ can be efficiently computed via nonuniform FFT (NUFFT) [155] or unequispaced FFT (USFFT) [156] methods. An overview of fast Fourier methods in the planar-wavefront SAR setting is given by Andersson et al. [157], and the spherical-wavefront case is presented in Demanet et al. [158]. Second, in

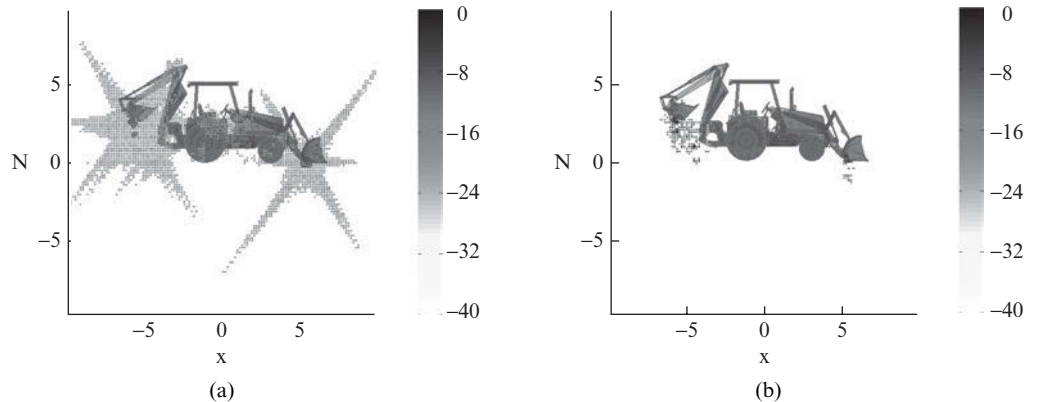


FIGURE 5-10 ■ (a) Backprojection and (b) QP_λ images, color scale is in dB.

the case of a rectilinear imaging grid, the operator $\mathbf{A}^H \mathbf{A}$ is a circulant matrix, meaning

$$\begin{aligned} (\mathbf{A}^H \mathbf{A})_{m,n} &\propto e^{i\xi_m \cdot (\mathbf{q}_m - \mathbf{q}_n)} \\ &= \sigma(m - n) \end{aligned} \quad (5.41)$$

Circulant matrices are diagonalizable by FFT matrices [20,159]. In our case $\mathbf{A}^H \mathbf{A} \mathbf{x}$ represents a convolution of the image \mathbf{x} with the PSF in the spatial domain. The product $\mathbf{A}^H \mathbf{A} \mathbf{x}$ can be implemented with FFTs by using the Fourier convolution identity. Specifically, if we write the PSF in the frequency domain as a diagonal matrix $\mathbf{\Sigma}_{psf}$ and the FFT matrix that transforms data from the spatial domain to the spatial-frequency domain as \mathbf{F} , then $\mathbf{A}^H \mathbf{A} \mathbf{x} = \mathbf{F}^H (\mathbf{\Sigma}_{psf} \mathbf{F} \mathbf{x})$. In this particular example, exploiting the circulant structure of $\mathbf{A}^H \mathbf{A}$ within each FISTA iteration amounts to a speed-up by about a factor of two over the NUFFT-only implementation.

5.4.2 Multipass 3-D Circular SAR

The second example in which sparsity can be exploited is wide-angle 3-D SAR imaging, in which angularly-dependent dictionary elements must be used due to the glint-like nature of certain target surfaces. Fundamentally the scene is sparse in that it consists of a two-dimensional surface embedded in \mathbb{R}^3 ; thus it seems reasonable to use this prior information to address the issue of undersampled data (the aircraft cannot fly enough passes to densely populate the viewing sphere). The data set we consider is the AFRL's publicly released *Volumetric SAR Data Set, Version 1.0* [160], which is also available from the AFRL VDL website [147]. The pulsed, monostatic, circular SAR collection system used a 9.6 GHz center-frequency (X band) LFM waveform with 640 MHz bandwidth, corresponding to a range resolution of 0.234 meters. The received data underwent the standard in-phase/quadrature (I/Q) demodulation, digitization, and match-filtering processes described in [161] and were digitally spotlighted [162] to a small area including a stationary civilian vehicle. The small size of the scene relative to the antenna location makes the planar-wavefront approximation extremely accurate.

The 360-degree angular diversity and multi elevational views included in this example may lead one to ask why regularization is required. The answer is that grating lobes of the 3-D PSF must be reduced. Despite significant elevational change between passes, the viewing sphere is undersampled in elevation, which leads to sever grating lobes along the direction orthogonal to the radar's range and cross-range directions for any given azimuthal subaperture. This effect is demonstrated in the backprojection image of Figure 5-11 (a).

Due to the shape of the target, we can expect its scattering behavior to have a strong view-angle dependence. One method for handling angular anisotropy is to introduce a point-scatterer fix by using a dictionary of point-like scatterers whose angular support is defined by a rectangular window in azimuth [163,164]. If the dictionary consists of disjoint subdictionaries (nonoverlapping subapertures), the objective function in QP_λ can be split into several pieces, that is

$$\arg \min_{\mathbf{x}} \|\mathbf{A} \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 = \arg \min_{\mathbf{x}} \sum_{k=1}^K \|\mathbf{A}_{\theta_k} \mathbf{x}_{\theta_k} - \mathbf{y}_{\theta_k}\|_2^2 + \lambda \|\mathbf{x}_{\theta_k}\|_1 \quad (5.42)$$

$$= \sum_{k=1}^K \left(\arg \min_{\mathbf{x}_{\theta_k}} \|\mathbf{A}_{\theta_k} \mathbf{x}_{\theta_k} - \mathbf{y}_{\theta_k}\|_2^2 + \lambda \|\mathbf{x}_{\theta_k}\|_1 \right) \quad (5.43)$$

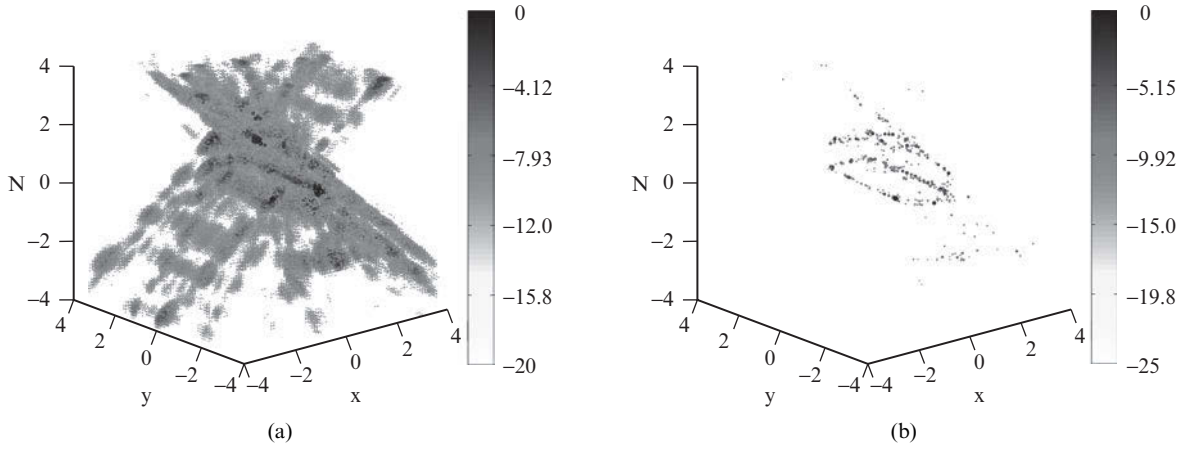


FIGURE 5-11 ■ (a) Backprojection and (b) Mixed MAP-ML images (dB scale).

Equation (5.43) shows that several smaller QP_λ problems may be solved in parallel—a big advantage considering the number of dictionary elements that were introduced; this embarrassingly-parallel formulation gains roughly two orders of magnitude in speed for realistic subaperture widths.

Given that a 3-D image is obtained for each azimuth window, we need to determine which value to display for each spatial pixel. This can be accomplished by modeling the pixels in each separate image (which we recall can be considered to be a MAP estimate)

$$I_k = \arg \min_{\mathbf{x}_{\theta_k}} \|\mathbf{A}_{\theta_k} \mathbf{x}_{\theta_k} - \mathbf{y}_{\theta_k}\|_2^2 + \lambda \|\mathbf{x}_{\theta_k}\|_1 \quad (5.44)$$

as independent test statistics for a maximum likelihood (ML) hypothesis test, that is, we set³⁸

$$I(x, y, z) = \max\{I_1, I_2, \dots, I_K\} \quad (5.45)$$

Due to the fact that the columns of \mathbf{A} have equal amplitudes, the ML scheme is equivalent to the generalized likelihood ratio test (GLRT). A description of the relationship between the GLRT and standard SAR imaging techniques is given in [166–168]. In this example we used 5-degree windows, which is a reasonable amount of expected persistence for vehicles at X band, within the hybrid MAP-ML scheme described by (5.44) and (5.45) to obtain the image in Figure 5-11. The number of data was $M = 1\,713\,743$ and the total image size was $N = (\text{number of subapertures})(\text{number of pixels}) = (72)(80^3) = 36,864,000$.

5.4.3 Multistatic Underground Imaging

Another radar imaging problem in which the underlying sparsity of the data gives rise to useful regularization approaches is GPR. Targets of interest such as underground

³⁸We should also note that this can be considered a crude way to enforce structure in the dictionary. An identical dictionary structure and approach can be applied to the moving-target imaging problem; see, for example [165].

tunnels are simple and have exploitable structure, such as well-defined sharp edges and low within-target contrast. In this example we will show how a priori knowledge of structured sparsity in the data allows us to appropriately regularize the otherwise ill-posed problem of underground imaging from an array of ultra narrowband antennas.

In all of our previous examples, we have assumed free-space propagation between the radar antennas and the scattering object while solving for a nondispersive target reflectivity function, $V(\mathbf{q}) = 1 - \frac{\epsilon(\mathbf{q})}{\epsilon_0}$ (this formulation necessarily assumes that the scattering material is isotropic and the materials have constant magnetic permeability μ_r that can be absorbed into the scattering reflectivity). In the case of GPR, the situation is complicated by the air-ground interface and possible frequency-dependent attenuation through the ground. In the context of the Born approximation, Deming [169] provides an algorithm that replaces the free-space Green's function with one that more accurately models propagation through lossy earth. Various formulations of the Green's function that also include the air-soil interface effects can be found in [170–173].

Assuming the Born approximation is reasonable and the direct-path contribution to the data has been removed, a generic linear model for the above-ground bistatic measured electric field (using dipole antennas of length ℓ) scattered from isotropic, nonmagnetic, homogeneous soil is [174]

$$E^{\text{scatt}}(\mathbf{y}_t, \mathbf{y}_r, \omega) = i\omega\mu_0k_0^2\ell \int \int \int \mathbf{b}_r^T \mathbf{G}(\mathbf{y}_r, \mathbf{q}, \omega) \cdot \underbrace{\mathbf{G}(\mathbf{q}, \mathbf{y}_t, \omega) \mathbf{a}_t}_{E^{\text{inc}}} V(\mathbf{q}) d^3\mathbf{q} \quad (5.46)$$

where we have introduced the scalar $k_0 = \omega/c_0$, the 3×3 dyadic Green's function \mathbf{G} , the transmit and receive locations $\mathbf{y}_t, \mathbf{y}_r \in \mathbb{R}^3$, and the vectors $\mathbf{a}_t, \mathbf{b}_r \in \mathbb{C}^3$ that encode the orientation, gain, and phase of the transmit and receive dipoles, respectively. If P transmitters simultaneously transmit, then the incident field is modeled as $E^{\text{inc}} = \sum_{p=1}^P \mathbf{G}(\mathbf{x}', \mathbf{y}_{t_p}, \omega) \mathbf{a}_{t_p}$, and the measured field (5.46) at the receiver becomes

$$E^{\text{scatt}}(\mathbf{y}_r, \omega) = i\omega\mu_0k_0^2\ell \int \int \int \mathbf{b}_r^T \mathbf{G}(\mathbf{y}_r, \mathbf{q}, \omega) \cdot \left(\sum_{p=1}^P \mathbf{G}(\mathbf{q}, \mathbf{y}_{t_p}, \omega) \mathbf{a}_{t_p} \right) V(\mathbf{q}) d^3\mathbf{q} \quad (5.47)$$

Equation (5.47) represents a linear equation that can be discretized to solve for the image V . By unwrapping the 3-D image V into the vector $\mathbf{x} \in \mathbb{C}^N$, concatenating the measurements of the scattered field $E^{\text{scatt}}(\mathbf{y}_r, \omega)$ into a vector $\mathbf{y} \in \mathbb{C}^M$ with $M = (\text{number of transmit frequencies})(\text{number of transmitter/receiver pairs})$, and discretizing the forward-operator $\mathbf{A} \in \mathbb{C}^{M \times N}$ into elements $A_{m,n}$ whose row number m indexes over all of the transmitter/receiver/frequency combinations and column number n indexes the voxels in the scene, we arrive at our usual ill-posed formulation $\mathbf{A}\mathbf{x} = \mathbf{y}$.

For our numerical example consider a synthetic transmitter/receiver setup in which the transmitters and receivers are arranged in a circular array placed on the air-soil interface 30 meters above an L-shaped tunnel (see Figure 5-12 for a top-down view of the sensor configuration and target). The forward data were simulated with the FDTD simulator GPRMAX [175] over the 4-7 MHz frequency range with 250 KHz steps. Although no noise was intentionally added to the data, the effects from invoking the Born approximation in the inversion algorithm, discretization errors, and imperfect removal of the direct-path signal were included. The number of measurements is $M = 360$, and the number of pixels is $N = 8405$.

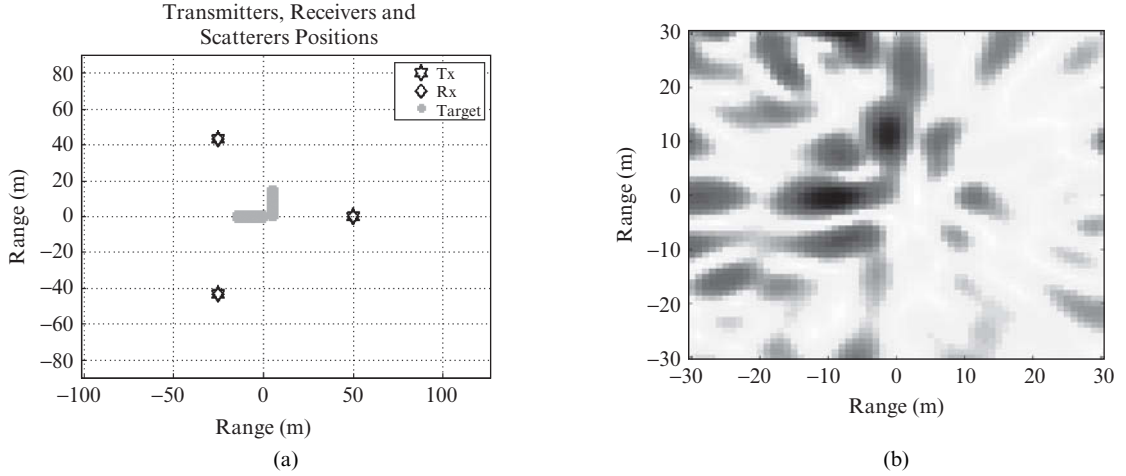


FIGURE 5-12 ■ (a) Sensor geometry and target and (b) a horizontal cross section of the Tikhonov-regularized image.

A common way to obtain an image \mathbf{x} when \mathbf{A} is ill-conditioned is to penalize large values of $\|\mathbf{x}\|_2$ by solving

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 + \alpha \|\mathbf{x}\|_2 \quad (5.48)$$

The tuning parameter α in the Tikhonov formulation (5.48) can be chosen by a variety of methods such as L-curve [69,176], generalized cross-validation [177], or Morozov's discrepancy principle [178]. Unfortunately, (5.48) fails to obtain a reasonable image (see Figure 5-12) because in this scenario, ℓ_2 penalization provides an insufficient amount of a priori information to effectively regularize the problem without destroying the quality of the solution. Fortunately for the tunnel-detection problem, our a priori knowledge also includes sparsity in both the image itself and its spatial gradient as well as physically motivated upper and lower bounds on the reflectivity.

We consider two different objective functions that seek to incorporate this knowledge. The first is

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 + \alpha \|\mathbf{x}\|_2 + \beta \|\mathbf{x}\|_1 \quad (5.49)$$

$$\text{s.t. } \tau_{\min} \leq |\mathbf{x}_n| \leq \tau_{\max}, \quad n = 1, \dots, N \quad (5.50)$$

To solve the problem (5.49)–(5.50) efficiently, we iteratively project solutions of (5.49) onto the feasible set (5.50). Decomposing the problem in this manner allows us to choose from several large-scale solvers (for example, LARS [179] was developed specifically for (5.49)). We choose to use the FISTA algorithm at each step by rewriting (5.49) and (5.50) as

$$\min_{\mathbf{x} \in \mathbb{C}^N} \left\| \begin{bmatrix} \mathbf{A} \\ \sqrt{\alpha} \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} \right\|_2 + \beta \|\mathbf{x}\|_1 \quad (5.51)$$

$$\text{s.t. } \tau_{\min} \leq |\mathbf{x}_n| \leq \tau_{\max}, \quad n = 1, \dots, N \quad (5.52)$$

which is the form of QP_λ .

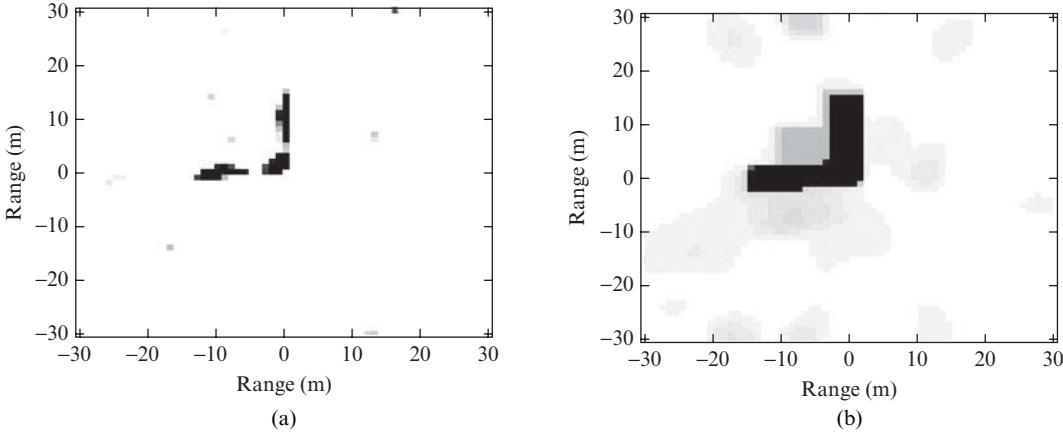


FIGURE 5-13 ■ (a) The ℓ_1/ℓ_2 -regularized image obtained solving (5.51) and (5.52) and (b) the $\ell_1/\ell_2/TV$ -regularized image obtained by solving (5.53) and (5.54). In both cases the regularization parameters (α, β) were chosen empirically.

Another potentially useful formulation is to also enforce sparsity in the spatial gradient. We define the problem

$$\min_{\mathbf{x} \in \mathbb{C}^N} \left\| \begin{bmatrix} \mathbf{A} \\ \sqrt{\alpha} \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} \right\|_2 + \alpha \|\mathbf{x}\|_{1,\epsilon} + \beta \|\mathbf{x}\|_{TV,\epsilon} \quad (5.53)$$

$$\text{s.t. } \tau_{\min} \leq |\mathbf{x}_n| \leq \tau_{\max}, \quad n = 1, \dots, N \quad (5.54)$$

where we have defined the smoothed norms (ϵ is a fixed small number)

$$\|\mathbf{x}\|_{1,\epsilon} = \sum_{n=1}^N \sqrt{|x_n|^2 + \epsilon} \quad (5.55)$$

$$\|\mathbf{x}\|_{TV,\epsilon} = \sum_{n=1}^{3(N_x-1)(N_y-1)(N_z-1)} \sqrt{[(\nabla_q |x|)_n]^2 + \epsilon} \quad (5.56)$$

where N_x, N_y, N_z denote the number of image pixels in the x, y, z directions, respectively, and the scalar $(\nabla_q |x|)_n$ is the n th element of the spatial gradient of the image amplitude. The definition (5.56) was chosen instead of the standard image gradient to allow for rapid phase oscillations which, as mentioned earlier in Section 5.3.1.2, should be expected in radar images. The second regularization term in (5.53) is a version of TV regularization, which enforces sparsity in the image's spatial gradient. The solution of (5.53) and (5.54) is solved by iteratively projecting unconstrained solutions of (5.53) onto the feasible set. The intermediate solutions to (5.53) were obtained with the algorithm described in [75]. See Figure 5-13 for the resulting images when using these modified objective functions.

5.4.4 A Herglotz Kernel Method

There exist radar scenarios in which the Born approximation is not necessarily valid. Without making the Born approximation, the standard radar imaging problem becomes nonlinear in the unknown variables (specifically, the unknown reflectivity and the unknown total electric field appear as a product of unknowns), as can be seen by examining

the Lippman-Schwinger integral equation for the scalar wave equation [180]. Here we illustrate one approach for addressing the multiple-scattering issue that is based on the LSM [181]. The exposition given here, which sketches the results in [182,183], considers the 2-D acoustic-scattering case. We note that this 2-D acoustic formulation can easily be extended to the 3-D electromagnetic case by using the analogous far-field relation used in [184,185].

The LSM solves a linear equation satisfied by the far-field data at a single frequency $\omega = kc$. Denote the far-field data for wavenumber k by $E_\infty(\hat{\gamma}(\theta_t), \hat{\gamma}(\theta_r), k)$, where $\hat{\gamma}(\theta_t) = [\cos \theta_t, \sin \theta_t]$ and $\hat{\gamma}(\theta_r) = [\cos \theta_r, \sin \theta_r]$ are the directions of the transmit and receive antennas, respectively, and assume we are imaging a target whose support is Γ . Then for a given incident plane wave from direction $\hat{\gamma}(\theta_t)$ and wavenumber k , the received data satisfy [186]

$$\int_0^{2\pi} E_\infty(\hat{\gamma}(\theta_t), \hat{\gamma}(\theta_r), k) g_p(\theta_r, k) d\theta_r = \frac{e^{j(\pi/4 - k\hat{\gamma}(\theta_t) \cdot \mathbf{p})}}{\sqrt{8\pi k}} \quad (5.57)$$

when $\mathbf{p} \in \mathbb{R}^2$ is within the scene. In (5.57), the unknown variable is the Herglotz kernel function, g_p .³⁹ Given a discrete number of antennas (assume all bistatic geometries are available), we arrive at the linear system $\mathbf{F}^{(k)} \mathbf{g}_p^{(k)} \approx \mathbf{v}^{(k)}(\mathbf{p})$, where the elements of $\mathbf{F}^{(k)}$ and $\mathbf{v}^{(k)}$ satisfy

$$\begin{aligned} \mathbf{F}_{m,n}^{(k)} &= E_\infty(\hat{\gamma}(\theta_m), \hat{\gamma}(\theta_n), k) \\ \mathbf{v}_m^{(k)}(\mathbf{p}) &= \frac{e^{j(\pi/4 - k\hat{\gamma}(\theta_m) \cdot \mathbf{p})}}{\sqrt{8\pi k}} \end{aligned}$$

The image is recovered by plotting the indicator function $\|\mathbf{g}_p^{(k)}\|_2^2$, which takes on large values for pixel location \mathbf{p} outside⁴⁰ Γ and is small inside Γ . Typically, each \mathbf{p} -dependent system of equations is solved in parallel by forming the block-diagonal system

$$\underbrace{\begin{bmatrix} \mathbf{F}^{(k)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^{(k)} & \dots & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{F}^{(k)} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{g}_{p_1}^{(k)} \\ \mathbf{g}_{p_2}^{(k)} \\ \vdots \\ \mathbf{g}_{p_P}^{(k)} \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \mathbf{v}^{(k)}(\mathbf{p}_1) \\ \mathbf{v}^{(k)}(\mathbf{p}_2) \\ \vdots \\ \mathbf{v}^{(k)}(\mathbf{p}_P) \end{bmatrix}}_{\mathbf{y}} \quad (5.58)$$

which we write as $\mathbf{Ax} = \mathbf{y}$ for brevity, where the number of rows of \mathbf{A} is $M = (\text{Number of pixels}) \times (\text{Number of transmitters}) = PT$ and the number of columns is $N = (\text{Number of pixels}) \times (\text{Number of receivers}) = PR$. Note that the far-field data composes the \mathbf{A} matrix, while \mathbf{y} is noiseless. The resulting type of least squares problem is referred to as a data least squares (DLS) problem [190]. Depending on the characteristics of the noise, specialized DLS techniques could be used. However, for convenience we ignore this subtlety and continue with the more common least-squares framework.

Because the validity of (5.58) depends on how well the receiving antennas populate the viewing sphere, it is clear that this method is very data-hungry and may require

³⁹The function g_p is known as the Herglotz kernel function because is the kernel of the Herglotz wave operator $(Hg)(\mathbf{p}) = \int_{\mathbb{S}} e^{ik\hat{\gamma} \cdot \mathbf{p}} g(\hat{\gamma}) d\hat{\gamma}$.

⁴⁰The reader may notice that this method is reminiscent of other imaging techniques such as the MUSIC algorithm [187]. The techniques are in fact related, and their similarities are discussed in [188,189].

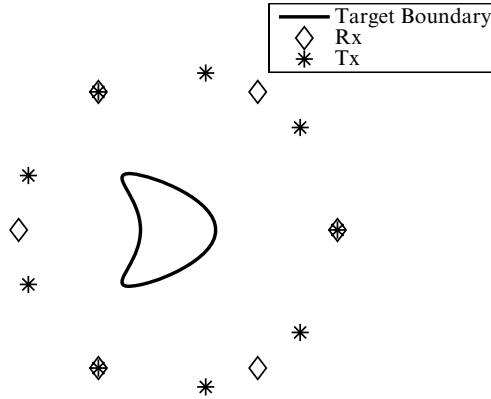


FIGURE 5-14 ■ The boundary of the kite target and the relative geometry of the antennas.

regularization methods. Much of the LSM-related literature uses Tikhonov regularization to solve (5.58). Here, we give an example in which a variant of TV regularization is used and compare its performance with Tikhonov regularization.

Defining the vector function $\mathbf{f}(\mathbf{x}) : \mathbb{C}^{PR} \mapsto \mathbb{R}^P$ whose n -th element is given by

$$f_n = \|\mathbf{x}_{p_n}\|_2^2 \quad (5.59)$$

we solve the TV problem [183]

$$\min_{\mathbf{x} \in \mathbb{C}^{PR}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\| + \lambda \|\mathbf{D}\mathbf{f}(\mathbf{x})\|_1 \quad (5.60)$$

where matrix \mathbf{D} computes the differences between adjacent pixel values f_n . For computational efficiency we solve (5.60) via the alternating minimization approach described in [191].

We consider the problem of imaging a kite-shaped target from an array of nine transmitters and six receivers. The target boundary and the directions of the transmit and receive antennas are given in Figure 5-14. The time-harmonic far-field data were synthesized via a combined single- and double-layer integral boundary approach along with a Nystrom method to discretize the integral equations [186]. The wavenumber was $k = 1$. The resulting Tikhonov- and TV-regularized reconstructions for the $\text{SNR} = 8$ dB case are given in Figures 5-15a and 5-15b, respectively. In this example, it is clear that the Tikhonov approach insufficiently regularizes the problem, while TV regularization provides a reasonable image. Due to the DLS form of the inverse problem described in (5.58), which uses measured data to populate the \mathbf{A} matrix, severely undersampled data introduce errors in the reconstructed image in a nonstandard way; instead of introducing grating lobes, we see the strange artifacts in Figure 5-15. This odd behavior is the result of the fact that a decrease in data yields a decrease in the number of Herglotz kernel function terms available to reconstitute the right-hand side of the equation, limiting our ability to effectively minimize the discrepancy term $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|$.

5.4.5 Single-Pulse Target Detection

In this section, we compute an example that illustrates the power of combining randomization with SR. Consider the problem of detecting a group of targets in delay (range) and Doppler using a single pulse. Naturally, the familiar radar ambiguity function encodes the information required for determining how well we can perform this task. A traditional

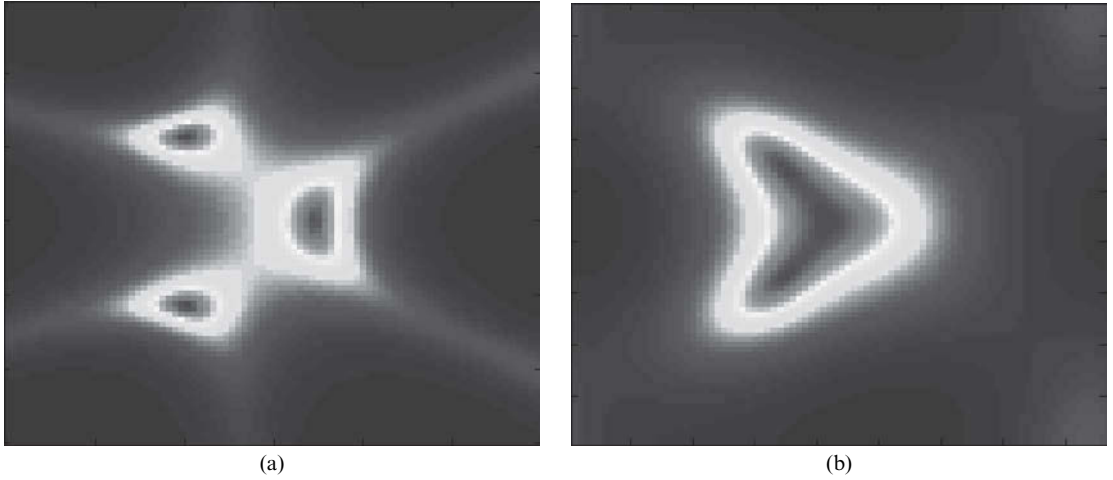


FIGURE 5-15 ■ (a) The Tikhonov-regularized and (b) the TV-regularized image obtained by solving (5.60). The Tikhonov regularization parameter was chosen by the maximum-curvature L-curve criterion, while the TV regularization parameter was chosen empirically.

chirp waveform has a ridge-like ambiguity function which leads to familiar issues with range-Doppler coupling [16]. Here we will instead consider comparison of a Gaussian pulse with a random, phase-only waveform. The Gaussian pulse is selected so that the widths of the ambiguity function main peak in delay and Doppler are approximately equal. The phase only waveform exhibits a thumbtack-like ambiguity function with a very sharp peak but relatively high and uniform sidelobes.

Our example is based heavily on the results in [56], where performance guarantees are proven for a cubic chirp waveform known as the Alltop sequence. The included proofs rely on the unique properties of Alltop sequences with prime length, but the authors point out that similar results hold for random phase waveforms. A more general discussion of CS forward operators obtained through random convolution can be found in [192]. We will choose a waveform of length 64 samples. Our dictionary will consist of all combinations of 64 shifts and 64 Doppler shifts of this waveform to yield $N = 64^2 = 4096$. The received signal has length $M = 64$, which represents 64 uniform samples of the received waveform.⁴¹ Analytical guarantees for reconstruction using this forward operator are provided in [56].

The results for this example are provided in Figure 5-16. The two rows show reconstructions using the Gaussian pulse (top) and the random phase waveform (bottom). The two images on the left provide matched filter results computed as $A^H y$, while the right column provides the solution⁴² to BP_σ with $\sigma = 0$. The Gaussian pulse results in a dictionary with a relatively high mutual coherence of 0.9851. In contrast, the noise waveform yields a thumbtack ambiguity function with mutual coherence of only 0.3205. The matched filter result is arguably better using the Gaussian pulse. The closely spaced targets are blurred together, but we at least obtain a rough idea of where in the delay-Doppler

⁴¹Note that we are assuming circular convolution with the transmitted pulse for simplicity.

⁴²These reconstructions are computed using SPGL1, but any solver could be used. Herman et al. [56] used the CVX package for their examples with similar results.

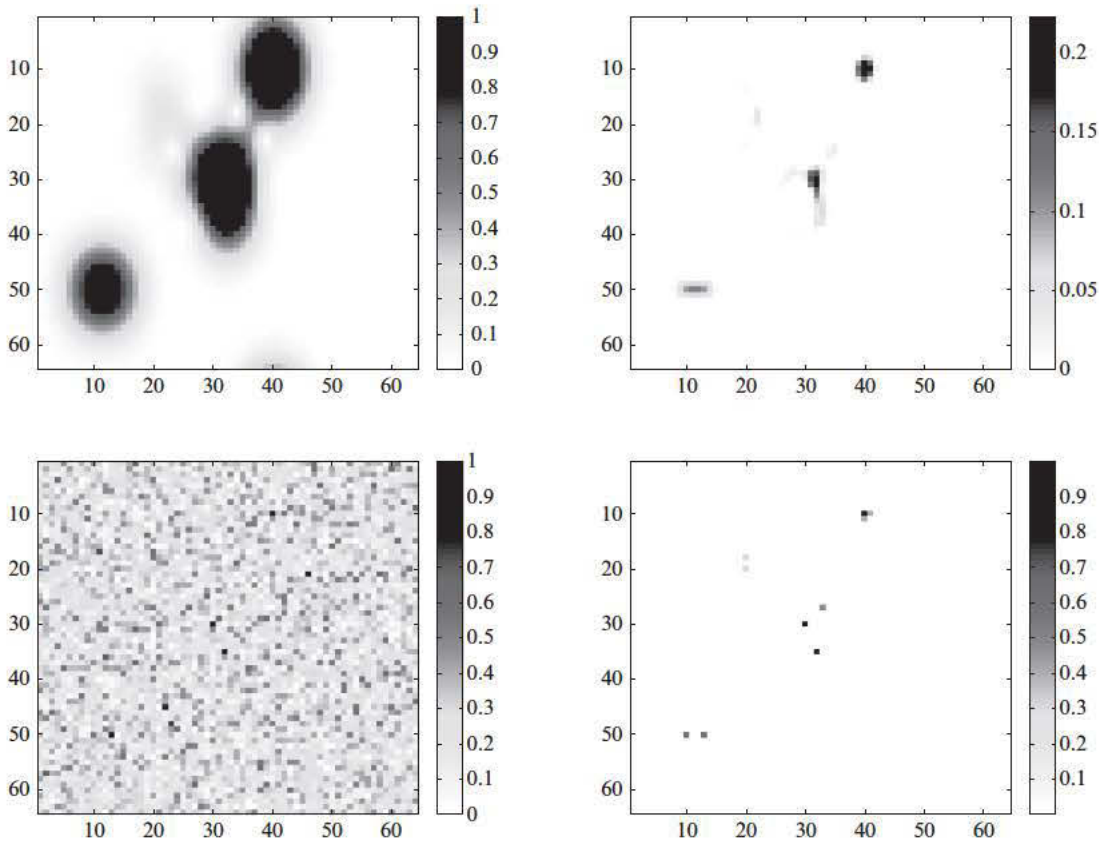


FIGURE 5-16 ■ Results for detecting 10 targets in delay-Doppler using a single pulse. Each plot shows the reconstructed reflectivity as a function of delay and Doppler bins, with delay on the x-axis. The top row shows results for a Gaussian pulse, while the bottom row is a random-phase, constant modulus signal. The left column provides matched filter results $A^H y$, while the right column shows BP_σ with $\sigma = 0$. The bottom right result matches the true signal almost perfectly.

plane the target energy is localized. The noise waveform's high sidelobes produce a very difficult to interpret result with matched filtering. Indeed, the two targets around delay bin 20 are totally submerged in the sidelobes from the stronger targets. We should emphasize that these simulations are noise-free; the interference is simply coherent addition of the sidelobes from the targets.

On the other hand, when we use an SR approach, the noise waveform results in a reconstruction that is essentially perfect as shown on the bottom right. There are actually three targets at delay bin 40 that are in adjacent delay-Doppler cells, and all three are accurately reconstructed. Even using the SR approach, the Gaussian pulse is not able to resolve these closely spaced targets. Thus, we clearly see that both SR (the right column) and randomization (the bottom row) must be combined to fully realize the benefits of CS. A related example is presented in [12] where a sequence of pulses from a group of cooperative radars is processed to form SAR images using the FOCUSS algorithm. Similar performance enhancements are observed when chirp waveforms are replaced with noise pulses.

5.5 | SUMMARY

We have shown that many common radar applications, even those that address multiple-scattering, dispersive media, and other complicated propagation effects, can be expressed in the standard linear framework for which SR and CS are applicable. Several examples were given in which the underlying scattering phenomenology dictates that the radar data are sparse in a known basis, motivating the use of SR and CS techniques. Although the use of SR algorithms in radar predates CS, the CS theory provides theoretical justification of their use by providing performance guarantees under well-defined measurement requirements such as the RIP and mutual coherence constraints. The measurement requirements themselves are often impossible to verify. However, because they are only sufficient conditions for guaranteed performance, in practice SR techniques work well even without strictly satisfying RIC or mutual coherence bounds. An added benefit of the CS theory is that the measurement requirements (which, when combined with SR, constitute CS) may guide the design of future radars and data acquisition strategies to collect better data for a given objective, as demonstrated by the waveform-design example given in the previous section. Beyond the standard sparsity that is now commonly exploited in the literature, structured sparsity appears in a variety of radar applications and dictionaries and, when incorporated into the SR algorithm, can be shown to significantly improve the performance guarantees. Advances in CS for structured, sparse data will undoubtedly play a role in future CS applications to radar.

Many radar applications are often limited by real-time requirements and large volumes of data, which until recently have made the application of SR to realistic scenarios merely a dream for practicing radar engineers. However, the recent academic interest in CS has rapidly accelerated research in SR techniques for large-scale problems, bringing many SR and CS applications within reach. Techniques for SR range from penalized least-squares techniques, which have excellent reconstruction performance at the expense of comparatively high computational burden, to iterative thresholding algorithms that have moderate computational cost while sometimes compromising their performance guarantees, to greedy approaches, which are the most computationally efficient but have the poorest performance guarantees (unless the greedy techniques are combined with structured-sparsity constraints, in which case they can outperform standard ℓ_1 solvers). Many of these techniques are grounded in a Bayesian setting and several can determine confidence levels on resulting reconstructions, which may prove extremely useful in decision tasks that rely on reconstructed images.

5.6 | FURTHER READING

A vast literature on CS has been published in the last several years, and new developments occur on a weekly or even daily basis. An excellent tutorial on CS ideas can be found in [21]. We recently coauthored [17], an article summarizing many of the same ideas contained in this chapter, along with additional historical references and more radar examples. Another recent survey of CS applications to radar, including examples on pulse compression, inverse SAR imaging, and direction of arrival estimation, can be found in [193].

Finally, as alluded to already, CS is still a rapidly evolving field. The best sources for up to date information on current trends and developments can be found online. The Rice

Compressed Sensing Resources list [132] includes links to tutorials, presentations, and papers on CS theory and applications. A specific section for CS radar is maintained. A slightly less formal although valuable reference can be found on the Nuit Blanche blog, which reports on developments in CS, provides articles on applications of CS ideas, and includes an occasionally updated discussion on “Compressive Sensing: The Big Picture,” which is a useful starting point in this research area. As already mentioned, an excellent list of algorithms is maintained in [60].

5.7 | ACKNOWLEDGMENTS

We would like to thank our many collaborators for their contributions to this chapter. In addition, we would like to thank Dr. Murali Rangaswamy, Edmund Zelnio, Professor Phillip Schniter, Professor Margaret Cheney, and the anonymous reviewers for insightful comments on the manuscript. J. Parker and M. Ferrara were supported under an AFOSR lab task under the direction of Dr. Arje Nachman. L. Potter was supported in part by AFOSR under Award FA9550-06-1-0324 and DARPA/ONR grant N66001-10-1-4090.

5.8 | REFERENCES

- [1] Candès, E., Romberg, J., and Tao, T., “Stable Signal Recovery from Incomplete and Inaccurate Measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [2] Donoho, D.L., “Compressed Sensing,” *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, April 2006.
- [3] Devaney, A.J., “Inverse Scattering within the Rytov Approximation,” *Optic Letters*, vol. 6, pp. 374–376, 1981.
- [4] Nolan, C.J. and Cheney, M., “Synthetic Aperture Inversion for Arbitrary Flight Paths and Nonflat Topography,” *IEEE Transactions on Image Processing*, vol. 12, pp. 1035–1043, September 2003.
- [5] Munson, Jr., D.C., O’Brien, J.D., and Jenkins, W.K., “A Tomographic Formulation of Spotlight-Mode Synthetic Aperture Radar,” *Proceedings of the IEEE*, vol. 71, pp. 917–925, August 1983.
- [6] Jakowatz, C.V., et al., *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Springer, 1996.
- [7] Morse, P.M. and Feshbach, H., *Methods of Theoretical Physics: Part I*, McGraw-Hill, New York, 1953.
- [8] Treves, F., *Basic Linear Partial Differential Equations*, Academic Press, New York, 1975.
- [9] Kak, A. and Slaney, M., *Principles of Computerized Tomographic Imaging*, IEEE, 1988.
- [10] Cheney, M. and Borden, B., *Fundamentals of Radar Imaging*, SIAM, Philadelphia, 2009.
- [11] Natterer, F., *The Mathematics of Computerized Tomography*, SIAM, Philadelphia, 2001.
- [12] Subotic, N.S., Thelen, B., Cooper, K., Buller, W., Parker, J., Browning, J., and Beyer, H., “Distributed Radar Waveform Design Based on Compressive Sensing Considerations,” *IEEE International Radar Conference*, 2008.
- [13] Skolnik, M., et al., *RADAR Handbook*, McGraw Hill, New York, 2008.

- [14] Li, J. and Stoica, P., *MIMO Radar Signal Processing*, Wiley, New York, 2009.
- [15] Melvin, W., "A STAP Overview," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1 (part 2), pp. 19–35, 2004.
- [16] Richards, M., *Fundamentals of Radar Signal Processing*, McGraw-Hill, New York, 2005.
- [17] Potter, L., Ertin, E., Parker, J., and Cetin, M., "Sparsity and Compressed Sensing in Radar Imaging," *Proceedings of the IEEE*, vol. 98, pp. 1006–1020, June 2010.
- [18] Khwaja, A.S. and Ma, J., "Applications of Compressed Sensing for SAR Moving-Target Velocity Estimation and Image Compression," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 8, pp. 2848–2860, 2011.
- [19] Candès, E. and Romberg, J., "Sparsity and Incoherence in Compressive Sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.
- [20] Golub, G. and Van Loan, C., *Matrix Computations*, Johns Hopkins University Press, 1996.
- [21] Candès, E.J. and Wakin, M.B., "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, March 2008.
- [22] Grant, M. and Boyd, S., "CVX: Matlab Software for Disciplined Convex Programming, Version 1.21," August 2010. Available at <http://cvxr.com/cvx>.
- [23] Tropp, J.A., "Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise," *IEEE Transactions on Information Theory*, vol. 52, pp. 1030–1051, March 2006.
- [24] Boyd, S. and Vandenberghe, L., *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
- [25] Taylor, H.L., Banks, S.C., and McCoy, J.F., "Deconvolution with the ℓ_1 norm," *Geophysics*, vol. 44, no. 1, pp. 39–52, 1979.
- [26] Tikhonov, A.N. and Arsenin, V.Y., *Solutions of Ill-Posed Problems*, Winston, Washington, DC, 1977.
- [27] Hansen, P.C., *Discrete Inverse Problems: Insight and Algorithms*, SIAM, Philadelphia, 2010.
- [28] Hoerl, A.E. and Kennard, R.W., "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, vol. 12, pp. 55–67, February 1970.
- [29] Baraniuk, R.G., Cevher, V., and Wakin, M.B., "Low-Dimensional Models for Dimensionality Reduction and Signal Recovery: A Geometric Perspective," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 959–971, 2010.
- [30] Parker, J.T. and Potter, L.C., "A Bayesian Perspective on Sparse Regularization for STAP Postprocessing," *IEEE International Radar Conference*, May 2010.
- [31] Chartrand, R. and Staneva, V., "Restricted Isometry Properties and Nonconvex Compressive Sensing," *Inverse Problems*, vol. 24, no. 3, p. 035020, 2008.
- [32] Kruskal, J., "Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics," *Linear Algebra and Its Applications*, vol. 18, no. 2, pp. 95–138, 1977.
- [33] Donoho, D. and Elad, M., "Optimally Sparse Representation in General (Nonorthogonal) Dictionaries via ℓ_1 Minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [34] Apostol, T.M., *Mathematical Analysis*, Addison-Wesley Publishing Company, Inc., 1974.
- [35] Candès, E., "The Restricted Isometry Property and Its Implications for Compressed Sensing," *Comptes rendus-Mathématique*, vol. 346, nos. 9–10, pp. 589–592, 2008.
- [36] S. Foucart, "A Note on Guaranteed Sparse Recovery via ℓ_1 Minimization," *Applied and Computational Harmonic Analysis*, 2009.

- [37] Austin, C.D., Moses, R.L., Ash, J.N., and Ertin, E., “On the Relation between Sparse Reconstruction and Parameter Estimation with Model Order Selection,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 560–570, 2010.
- [38] Candès, E., Eldar, Y., and Needell, D., “Compressed Sensing with Coherent and Redundant Dictionaries,” *Arxiv preprint arXiv*, vol. 1005, no. 2613, 2010.
- [39] Needell, D. and Tropp, J.A., “CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples,” *Applied Computational Harmonic Analysis*, vol. 26, pp. 301–321, 2008.
- [40] Iouditski, A. and Nemirovski, A., “On Verifiable Sufficient Conditions for Sparse Signal Recovery via ℓ_1 Minimization,” *Preprint*, 2010.
- [41] Bickel, P., Ritov, Y., and Tsybakov, A., “Simultaneous analysis of Lasso and Dantzig selector,” *Annals of Statistics*, vol. 37, no. 4, pp. 1705–1732, 2009.
- [42] Xu, W. and Hassibi, B., “On Sharp Performance Bounds for Robust Sparse Signal Recoveries,” in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 493–497, 2009.
- [43] Howard, S., Calderbank, A., and Searle, s., “A Fast Reconstruction Algorithm for Deterministic Compressive Sensing Using Second Order Reed-Muller Codes,” in *42nd Annual Conference on Information Sciences and Systems*, pp. 11–15, 2008.
- [44] Jafarpour, S., Xu, W., Hassibi, B., and Calderbank, R., “Efficient and Robust Compressed Sensing Using Optimized Expander Graphs,” *IEEE Transactions on Information Theory*, vol. 55, pp. 4299–4308, September 2009.
- [45] Rudelson, M. and Vershynin, R., “On Sparse Reconstruction from Fourier and Gaussian Measurements,” *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.
- [46] Baraniuk, R., Davenport, M., DeVore, R., and Wakin, M., “A Simple Proof of the Restricted Isometry Property for Random Matrices,” *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [47] DeVore, R., “Deterministic Constructions of Compressed Sensing Matrices,” *Journal of Complexity*, vol. 23, nos. 4–6, pp. 918–925, 2007.
- [48] Lo, Y.T., “A Mathematical Theory of Antenna Arrays with Randomly Spaced Elements,” *IEEE Transactions on Antennas and Propagation*, vol. 12, pp. 257–268, 1964.
- [49] Unz, H., “Linear Arrays with Arbitrarily Distributed Elements,” *IRE Transactions on Antennas and Propagation*, vol. 8, pp. 222–223, March 1960.
- [50] Harrington, R.F., “Sidelobe Reduction by Nonuniform Element Spacing,” *IRE Transactions on Antennas and Propagation*, vol. 9, pp. 187–192, March 1961.
- [51] Axelsson, S., “Noise Radar Using Random Phase and Frequency Modulation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 2370–2384, November 2004.
- [52] Horton, B., “Noise-Modulated Distance Measuring Systems,” *Proceedings of the IRE*, vol. 47, pp. 821–828, May 1959.
- [53] Donoho, D.L., Elad, M., and Temlyakov, V.N., “Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [54] Ben-Haim, Z., Eldar, Y.C., and Elad, M., “Coherence-Based Performance Guarantees for Estimating a Sparse Vector under Random Noise,” *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5030–5043, 2010.
- [55] Blanchard, J.D., Cartis, C., and Tanner, J., “Decay Properties of Restricted Isometry Constants,” *IEEE Signal Processing Letters*, vol. 16, pp. 572–575, July 2009.

- [56] Herman, M.A. and Strohmer, T., “High-Resolution Radar via Compressed Sensing,” *IEEE Transactions on Signal Processing*, vol. 57, pp. 2275–2284, June 2009.
- [57] Levanon, N. and Mozeson, E., *Radar Signals*, John Wiley & Sons, Interscience Division, New York, 2004.
- [58] Cociorva, P., “Preconditioners for Sparse Linear Inversion,” Master’s thesis, The Ohio State University, 2006.
- [59] Duarte, M. and Baraniuk, R., “Spectral Compressive Sensing,” *Preprint*, 2010.
- [60] Becker, S., “List of Sparse and Low-Rank Recovery Algorithms.” Available at <http://www.ugcs.caltech.edu/~srbecker/algorithms.shtml>.
- [61] Becker, S., Bobin, J., and Candès, E., “NESTA: A Fast and Accurate First-Order Method for Sparse Recovery,” *SIAM Journal on Imaging Sciences*, vol. 4, p. 1, 2011.
- [62] Burns, J.W., Subotic, N.S., and Pandelis, D., “Adaptive Decomposition in Electromagnetics,” in *Antennas and Propagation Society International Symposium*, pp. 1984–1987, July 1997.
- [63] van den Berg, E. and Friedlander, M., “Probing the Pareto Frontier for Basis Pursuit Solutions,” *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.
- [64] Tibshirani, R., “Regression Shrinkage and Selection via the LASSO,” *Journal of the Royal Statistical Society, Series B (Methodological)*, pp. 267–288, 1996.
- [65] Candès, E. and Tao, T., “The Dantzig Selector: Statistical Estimation When p Is Much Larger than n ,” *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.
- [66] Donoho, D., Maleki, A., and Montanari, A., “Message-Passing Algorithms for Compressed Sensing,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.
- [67] Osborne, M., Presnell, B., and Turlach, B., “On the LASSO and Its Dual,” *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 319–337, 2000.
- [68] Lawson, C.L. and Hanson, R.J., *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
- [69] Hansen, P.C., “Analysis of Discrete Ill-Posed Problems by Means of the L-Curve,” *SIAM Review*, vol. 34, no. 4, pp. 561–580, 1992.
- [70] Chen, S., Donoho, D., and Saunders, M., “Atomic Decomposition by Basis Pursuit,” *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
- [71] Delaney, A.H. and Bresler, Y., “Globally Convergent Edge-Preserving Regularized Reconstruction: An Application to Limited-Angle Tomography,” *IEEE Transactions on Image Processing*, vol. 7, no. 2, pp. 204–221, 1998.
- [72] Lustig, M., Donoho, D., and Pauly, J., “Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, p. 1182, 2007.
- [73] Smith, L.M., Keegan, M.S., Wittman, T., Mohler, G.O., and Bertozzi, A.L., “Improving Density Estimation by Incorporating Spatial Information,” *EURASIP Journal on Advances in Signal Processing*, 2010.
- [74] van den Berg, E. and Friedlander, M., “Sparse Optimization with Least-Squares Constraints,” Technical Report TR-2010-02, Department of Computer Science, University of British Columbia, 2010.
- [75] Cetin, M. and Karl, W.C., “Feature-Enhanced Synthetic Aperture Radar Image Formation Based on Nonquadratic Regularization,” *IEEE Transactions on Image Processing*, vol. 10, pp. 623–631, April 2001.
- [76] Rudin, L., Osher, S., and Fatemi, E., “Nonlinear Total Variation Based Noise Removal Algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

- [77] Kragh, T.J. and Kharbouch, A.A., “Monotonic Iterative Algorithms for SAR Image Restoration,” in *Proceedings of the IEEE International Conference on Image Processing*, pp. 645–648, October 8–11, 2006.
- [78] Hunter, D. and Lange, K., “A Tutorial on MM Algorithms,” *American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.
- [79] Beck, A. and Teboulle, M., “A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [80] Blumensath, T. and Davies, M.E., “Iterative Hard Thresholding for Compressed Sensing,” *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [81] Cevher, V., “An ALPS View of Sparse Recovery,” in *Proceedings of ICASSP*, pp. 5808–5811, May 2011.
- [82] Lawson, C.L., “Contributions to the Theory of Linear Least Maximum Approximations,” PhD thesis, University of California–Los Angeles, 1961.
- [83] Lee, H., Sullivan, D., and Huang, T., “Improvement of Discrete Band-Limited Signal Extrapolation by Iterative Subspace Modification,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1569–1572, 1987.
- [84] Brito, A.E., Chan, S.H., and Cabrera, S.D., “SAR Image Formation Using 2-D Reweighted Minimum Norm Extrapolation,” *SPIE*, vol. 3721, pp. 78–91, 1999.
- [85] Daubechies, I., DeVore, R., Fornasier, M., and Gunturk, S., “Iteratively Re-weighted Least Squares Minimization for Sparse Recovery,” *Communications on Pure and Applied Mathematics*, vol. 14, pp. 877–905, December 2009.
- [86] Gorodnitsky, I.F. and Rao, B.D., “Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm,” *IEEE Transactions on Signal Processing*, vol. 45, pp. 600–616, March 1997.
- [87] Candès, E.J., Wakin, M., and Boyd, S., “Enhancing Sparsity by Reweighted ℓ_1 Minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2007.
- [88] Chartrand, R. and Yin, W., “Iteratively Reweighted Algorithms for Compressive Sensing,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 3869–3872, 2008.
- [89] Høgbom, J., “Aperture Synthesis with a Non-regular Distribution of Interferometer Baselines,” *Astrophysics and Astrophysics Supplement Series*, vol. 15, pp. 417–426, 1974.
- [90] Mallat, S.G. and Zhang, Z., “Matching Pursuits with Time-Frequency Dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, December 1993.
- [91] Tropp, J.A. and Gilbert, A.C., “Signal Recovery from Random Measurements via Orthogonal Matching Pursuit,” *IEEE Transactions on Information Theory*, vol. 53, pp. 4655–4666, December 2007.
- [92] Tropp, J.A., “Greed Is Good: Algorithmic Results for Sparse Approximation,” *IEEE Transactions on Information Theory*, vol. 50, pp. 2231–2242, October 2004.
- [93] Dai, W. and Milenkovic, O., “Subspace Pursuit for Compressive Sensing Signal Reconstruction,” *IEEE Transactions on Information Theory*, vol. 55, pp. 2230–2249, May 2009.
- [94] Foucart, S., “Sparse Recovery Algorithms: Sufficient Conditions in Terms of Restricted Isometry Constants,” *Preprint*, 2010.
- [95] Maleki, A. and Donoho, D.L., “Optimally Tuned Iterative Reconstruction Algorithms for Compressed Sensing,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 330–341, 2010.

- [96] Elad, M. and Yavneh, I., “A Plurality of Sparse Representations Is Better than the Sparsest One Alone,” *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4701–4714, 2009.
- [97] Schniter, P., Potter, L., and Ziniel, J., “Fast Bayesian Matching Pursuit,” in *Proceedings of the Workshop on Information Theory and Applications*, La Jolla, CA, January 2008.
- [98] Baron, D., Sarvotham, S., and Baraniuk, R., “Bayesian Compressive Sensing via Belief Propagation,” *IEEE Transactions on Signal Processing*, vol. 58, pp. 269–280, January 2010.
- [99] Rangan, S., “Estimation with Random Linear Mixing, Belief Propagation and Compressed Sensing,” *Arxiv preprint arXiv:1001.2228*, 2010.
- [100] Frey, B., *Graphical Models for Machine Learning and Digital Communication*, MIT Press, Cambridge, MA, 1998.
- [101] Montanari, A., “Graphical Models Concepts in Compressed Sensing,” Submitted, 2010.
- [102] Donoho, D. and Tanner, J., “Observed Universality of Phase Transitions in High-Dimensional Geometry, with Implications for Modern Data Analysis and Signal Processing,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1906, p. 4273, 2009.
- [103] Donoho, D., Maleki, A., and Montanari, A., “The Noise-Sensitivity Phase Transition in Compressed Sensing,” *Arxiv preprint arXiv:1004.1218*, 2010.
- [104] Donoho, D., Maleki, A., and Montanari, A., “Message Passing Algorithms for Compressed Sensing: I—Motivation and Construction,” in *Proceedings of Information Theory Workshop*, 2010.
- [105] Donoho, D., Maleki, A., and Montanari, A., “Message Passing Algorithms for Compressed Sensing: II—Analysis and Validation,” in *Proceedings of Information Theory Workshop*, 2010.
- [106] Maleki, A. and Montanari, A., “Analysis of Approximate Message Passing Algorithm,” in *Proceedings of the 44th Annual Conference on Information Sciences and Systems*, pp. 1–7, 2010.
- [107] Bayati, M. and Montanari, A., “The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing,” *Arxiv preprint arXiv:1001.3448*, 2010.
- [108] Rangan, S., “Generalized Approximate Message Passing for Estimation with Random Linear Mixing,” *Arxiv preprint arXiv:1010.5141*, 2010.
- [109] Proakis, J. and Salehi, M., *Digital Communications*, McGraw-Hill, New York, 1995.
- [110] Amin Khajehnejad, M., Xu, W., Salman Avestimehr, A., and Hassibi, B., “Weighted ℓ_1 minimization for sparse recovery with prior information,” in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 483–487, 2009.
- [111] Cevher, V., Duarte, M., Hegde, C., and Baraniuk, R., “Sparse Signal Recovery Using Markov Random Fields,” in *Proceedings of the Workshop on Neural Information Processing Systems*, 2008.
- [112] Baraniuk, R.G., Cevher, V., Duarte, M.F., and Hegde, C., “Model-Based Compressive Sensing,” *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [113] Cevher, V., Indyk, P., Carin, L., and Baraniuk, R.G., “Sparse Signal Recovery and Acquisition with Graphical Models,” *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 92–103, 2010.
- [114] Eldar, Y., Kuppinger, P., and Bölcskei, H., “Compressed Sensing of Block-Sparse Signals: Uncertainty Relations and Efficient Recovery,” *Preprint arXiv: 0906.3173*, 2009.
- [115] Ramakrishnan, N., Ertin, E., and Moses, R.L., “Joint Enhancement of Multichannel SAR Data,” in *Algorithms for Synthetic Aperture Radar Imagery XIV*, vol. 6568, SPIE, April 2007.

- [116] Eldar, Y.C. and Mishali, M., “Robust Recovery of Signals from a Structured Union of Subspaces,” *IEEE Transactions on Information Theory*, vol. 55, pp. 5302–5316, November 2009.
- [117] Schniter, P., “Turbo Reconstruction of Structured Sparse Signals,” *ICASSP*, 2010.
- [118] Koetter, R., Singer, A., and Tuchler, M., “Turbo Equalization,” *IEEE Signal Processing Magazine*, vol. 21, pp. 67–80, January 2004.
- [119] Candès, E. and Recht, B., “Exact Matrix Completion via Convex Optimization,” *Arxiv preprint arXiv:0805.44*, 2008.
- [120] Candès, E., Li, X., Ma, Y., and Wright, J., “Robust Principal Component Analysis?: Recovering Low-Rank Matrices from Sparse Errors,” in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 201–204, 2010.
- [121] Tao, M. and Yuan, X., “Recovering Low-Rank and Sparse Components of Matrices from Incomplete and Noisy Observations,” *SIAM Journal on Optimization*, vol. 21, no. 1, pp. 57–81, 2011.
- [122] Cao, G. and Bouman, C.A., “Covariance Estimation for High Dimensional Data Vectors Using the Sparse Matrix Transform,” in *Proceedings of Neural Information Processing Systems Conference*, December 2008.
- [123] Malioutov, D., Cetin, M., and Willsky, A.S., “A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays,” *IEEE Transactions on Signal Processing*, vol. 53, pp. 3010–3022, August 2005.
- [124] Chi, Y., Pezeshki, A., Scharf, L., and Calderbank, R., “Sensitivity to Basis Mismatch in Compressed Sensing,” in *Proceedings of the IEEE International Acoustics Speech and Signal Processing Conference*, pp. 3930–3933, 2010.
- [125] Zhu, H., Leus, G., and Giannakis, G., “Sparsity-Cognizant Total Least-Squares for Perturbed Compressive Sampling,” *IEEE Transactions on Signal Processing*, vol. 59, pp. 2002–2016, May 2011.
- [126] Herman, M.A. and Strohmer, T., “General Deviants: An Analysis of Perturbations in Compressed Sensing,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, 2010.
- [127] Rosenbaum, M. and Tsybakov, A., “Sparse Recovery under Matrix Uncertainty,” *Annals of Statistics*, vol. 38, no. 5, pp. 2620–2651, 2010.
- [128] Parker, J.T., Cevher, V., and Schniter, P., “Compressive Sensing under Matrix Uncertainties: An Approximate Message Passing Approach,” in *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, November 2011.
- [129] Tomic, I. and Frossard, P., “Dictionary Learning,” *IEEE Signal Processing Magazine*, vol. 28, pp. 27–38, March 2011.
- [130] Goldstein, T. and Osher, S., “The Split Bregman Method for ℓ_1 -Regularized Problems,” *SIAM Journal on Imaging Sciences*, vol. 2, p. 323, 2009.
- [131] Yin, W., Osher, S., Goldfarb, D., and Darbon, J., “Bregman Iterative Algorithms for ℓ_1 -Minimization with Applications to Compressed Sensing,” *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 143–168, 2008.
- [132] Rice University, “Compressive Sensing Resources.” Available at <http://dsp.rice.edu/cs>.
- [133] Gurbuz, A.C., McClellan, J.H., and Scott Jr., R., “A Compressive Sensing Data Acquisition and Imaging Method for Stepped Frequency GPRs,” *IEEE Transactions on Signal Processing*, pp. 1–1, 2009.

- [134] Bhattacharya, S., Blumensath, T., Mulgrew, B., and Davies, M., “Fast Encoding of Synthetic Aperture Radar Raw Data Using Compressed Sensing,” in *Proceedings of the IEEE/SP 14th Workshop on Statistical Signal Processing*, pp. 448–452, August 26–29, 2007.
- [135] Novak, L., “The Effects of SAR Data Compression on Coherent and Non-coherent Change Detection,” in *IEEE International Radar Conference*, 2009.
- [136] Patel, V.M., Easley, G.R., Dennis, J. Healy, M., and Chellappa, R., “Compressed Synthetic Aperture Radar,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, pp. 244–254, April 2010.
- [137] Baraniuk, R. and Steeghs, P., “Compressive Radar Imaging,” in *Proceedings of the IEEE Radar Conference*, pp. 128–133, April 17–20, 2007.
- [138] Romberg, J., “Compressive Sensing by Random Convolution,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.
- [139] Eldar, Y., “Compressed Sensing of Analog Signals in Shift-Invariant Spaces,” *IEEE Transactions on Signal Processing*, vol. 57, pp. 2986–2997, August 2009.
- [140] Ross, T., Worrell, S., Velten, V., Mossing, J., and Bryant, M., “Standard SAR ATR Evaluation Experiments Using the MSTAR Public Release Data Set,” in *Proceedings of the SPIE: Algorithms for Synthetic Aperture Radar Imagery V*, Orlando, FL, pp. 566–573, 1998.
- [141] Varshney, K.R., Cetin, M., Fisher, J.W., and Willsky, A.S., “Sparse Representation in Structured Dictionaries with Application to Synthetic Aperture Radar,” *IEEE Transactions on Signal Processing*, vol. 56, pp. 3548–3561, August 2008.
- [142] Ji, S., Xue, Y., and Carin, L., “Bayesian Compressive Sensing,” *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [143] Angelosante, D., Giannakis, G.B., and Grossi, E., “Compressed Sensing of Time-Varying Signals,” in *International Conference on Digital Signal Processing*, pp. 1–8, July 2009.
- [144] Vaswani, N. and Lu, W., “Modified-CS: Modifying compressive sensing for problems with partially known support,” *IEEE Transactions on Signal Processing*, vol. 58, pp. 4595–4607, September 2010.
- [145] Ziniel, J., Potter, L.C., and Schniter, P., “Tracking and Smoothing of Time-Varying Sparse Signals via Approximate Belief Propagation,” in *Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, November 2010.
- [146] Naidu, K. and Lin, L., “Data Dome: Full k-Space Sampling Data for High Frequency Radar Research,” Pp. 200–207 in *Algorithms for Synthetic Aperture Radar Imagery XI*, vol. 5427, Ed. E.G. Zelnio and F.D. Garber, SPIE, 2004.
- [147] Air Force Research Laboratory, “Virtual Distributed Laboratory.” Available at <https://www.vdl.afrl.af.mil/>.
- [148] Stuff, M.A., “Three-Dimensional Analysis of Moving Target Radar Signals: Methods and implications for ATR and feature aided tracking,” pp. 485–496 in *Algorithms for Synthetic Aperture Radar Imagery VI*, vol. 3721, Ed. E.G. Zelnio, Proceedings of SPIE, 1999.
- [149] Stuff, M.A., Sanchez, P., and Biancalana, M., “Extraction of Three-Dimensional Motion and Geometric Invariants from Range Dependent Signals,” *Multidimensional Systems of Signal Processing*, vol. 14, nos. 1–3, pp. 161–181, 2003.
- [150] Keller, J.B., “Geometrical Theory of Diffraction,” *Journal of the Optical Society of America*, pp. 116–130, January 1962.
- [151] Potter, L.C., Chiang, D., Carriere, R., and Gerry, M.J., “A GTD-Based Parametric Model for Radar Scattering,” *IEEE Transactions on Antennas and Propagation*, vol. 43, no. 10, pp. 1058–1067, 1995.

- [152] Rigling, B.D. and Moses, R.L., "GTD-Based Scattering Models for Bistatic SAR," pp. 208–219 in *Algorithms for Synthetic Aperture Radar Imagery XI*, vol. 5427, Ed. E.G. Zelnio, SPIE, 2004.
- [153] Jackson, J., Rigling, B., and Moses, R., "Canonical Scattering Feature Models for 3-D and Bistatic SAR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 2, pp. 525–541, 2010.
- [154] Jackson, J.A. and Moses, R.L., "An Algorithm for 3-D Target Scatterer Feature Estimation from Sparse SAR Apertures," in *Algorithms for Synthetic Aperture Radar Imagery XVI*, vol. 7337, Ed. E.G. Zelnio and F.D. Garber, Proceedings of SPIE, 2009.
- [155] Greengard, L. and Lee, J.Y., "Accelerating the Nonuniform Fast Fourier Transform," *SIAM Review*, vol. 46, pp. 443–454, 2004.
- [156] Beylkin, G., Gorman, J., Li-Fliss, S., and Ricoy, M., "SAR Imaging and Multiresolution Analysis," pp. 2144–2152 in *Proceedings of the SPIE: Algorithms for Synthetic Aperture Radar Imagery II*, vol. 2487, Ed. D.A. Giglio, pp. 2144–152, SPIE, 1995.
- [157] Andersson, F., Moses, R., and Natterer, F., "Fast Fourier Methods for Synthetic Aperture Radar Imaging," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, January 2012.
- [158] Demanet, L., Ferrara, M., Maxwell, N., Poulson, J., and Ying, L., "A Buttery Algorithm for Synthetic Aperture Radar Imaging," Submitted to *SIAM Imaging Sciences*, September 2010.
- [159] Delaney, A. and Bresler, Y., "A Fast and Accurate Fourier Algorithm for Iterative Parallel-Beam Tomography," *IEEE Transactions on Image Processing*, vol. 5, pp. 740–753, May 1996.
- [160] Curtis, J., Casteel, H., Gorham, L.A., Minardi, M.J., Scarborough, S.M., Naidu, K.D., and Majumder, U.K., "A Challenge Problem for 2-D/3-D Imaging of Targets from a Volumetric Data Set in an Urban Environment," P. 65680D in *Proceedings of SPIE*, vol. 6568, Ed. E.G. Zelnio and F.D. Garber, SPIE, 2007.
- [161] Wahl, D.E., Eichel, P.H., Ghiglia, D.C., Thompson, P.A., and Jakowatz, C.V., *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Springer, 1996.
- [162] Soumekh, M., *Synthetic Aperture Radar Signal Processing with MATLAB Algorithms*, John Wiley & Sons, Interscience Division, New York, 1999.
- [163] Moses, R.L. and Potter, L.C., "Noncoherent 2-D and 3-D SAR Reconstruction from Wide-Angle Measurements," in *13th Annual Adaptive Sensor Array Processing Workshop*, MIT Lincoln Laboratory, 2005.
- [164] Austin, C.D., Ertin, E., and Moses, R.L., "Sparse Multipass 3-D SAR Imaging: Applications to the GOTCHA Data Sets," in *Algorithms for Synthetic Aperture Radar Imagery XVI*, Ed. E.G. Zelnio and F.D. Garber, Orlando, FL, SPIE Defense and Security Symposium, April 13–17, 2009.
- [165] Stojanovic, I. and Karl, W., "Imaging of Moving Targets with Multi-static SAR Using an Overcomplete Dictionary," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 164–176, 2010.
- [166] Rossi, D.J. and Willsky, A.S., "Reconstruction from Projections Based on Detection and Estimation of Objects—Part I: Performance Analysis and Part II: Robustness Analysis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, pp. 886–906, August 1984.
- [167] Ertin, E., Potter, L., and Moses, R., "Enhanced Imaging over Complete Circular Apertures," in *40th Asilomar Conference on Signals, Systems and Computers*, pp. 1580–1584, October 29–November 1, 2006.

- [168] Voccola, K., Yazici, B., Ferrara, M., and Cheney, M., “On the Relationship between the Generalized Likelihood Ratio Test and Backprojection for Synthetic Aperture Radar Imaging,” in *Automatic Target Recognition XIX*, vol. 7335, Ed. F.A. Sadjadi and A. Mahalanobis, Proceedings of the SPIE, 2009.
- [169] Deming, R.W. and Devaney, A.J., “Diffraction Tomography for Multi-monostatic Ground Penetrating Radar Imaging,” *Inverse Problems*, vol. 13, no. 1, p. 29, 1997.
- [170] P. Meincke, “Linear GPR Inversion for Lossy Soil and a Planar Air-Soil Interface,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 2713–2721, December 2001.
- [171] Cui, T.J. and Chew, W.J., “Diffraction Tomographic Algorithm for the Detection of Three-Dimensional Objects Buried in a Lossy Half-Space,” *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 1, pp. 42–49, 2002.
- [172] Zhadanov, M.S., *Geophysical Inverse Theory and Regularization Problems*, Elsevier Science Inc., 2002.
- [173] Lo Monte, L., Erricolo, D., Soldovieri, F., and Wicks, M.C., “Radio Frequency Tomography for Tunnel Detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, pp. 1128–1137, March 2010.
- [174] Lo Monte, L. and Parker, J., “Sparse Reconstruction Methods in RF Tomography for Underground Imaging,” in *International Waveform Design and Diversity Conference*, 2010.
- [175] Giannopoulos, A., “GPRMAX Simulator.” Available at <http://www.gprmax.org>.
- [176] Hansen, P.C., “The L-Curve and Its Use in the Numerical Treatment of Inverse Problems,” in *Computational Inverse Problems in Electrocardiology*, Ed. P. Johnston, pp. 119–142, WIT Press, Southampton, 2001.
- [177] Golub, G.H. and von Matt, U., “Generalized Cross-Validation for Large Scale Problems,” *J. Comput. Graph. Stat.*, vol. 6, pp. 1–34, 1995.
- [178] Scherzer, O., “The Use of Morozov’s Discrepancy Principle for Tikhonov Regularization for Solving Nonlinear Ill-Posed Problems,” *Computing*, vol. 51, pp. 45–60, 1993. 10.1007/BF02243828.
- [179] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R., “Least Angle Regression,” *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [180] Cheney, M. and Borden, B., “Problems in Synthetic-Aperture Radar Imaging,” *Inverse Problems*, vol. 25, no. 12, p. 123005, 2009.
- [181] Colton, D. and Kirsch, A., “A Simple Method for Solving Inverse Scattering Problems in the Resonance Region,” *Inverse Problems*, vol. 12, no. 4, p. 383, 1996.
- [182] Alqadah, H., Parker, J., Ferrara, M., and Fan, H., “A Single Frequency Sparsity-Constrained Linear Sampling Method,” in *Defense Applications of Signal Processing*, 2009.
- [183] Alqadah, H., Fan, H., Ferrara, M., and Parker, J., “A Modified Total-Variation Approach for Single-Frequency Inverse Scattering,” in *IEEE Asilomar Conference on Signals, Systems, and Computers*, 2010.
- [184] Potthast, R., *Point Sources and Multipoles in Inverse Scattering Theory*, Chapman and Hall/CRC, 2001.
- [185] Brignone, M., Bozza, G., Aramini, R., Pastorino, M., and Piana, M., “A Fully No-Sampling Formulation of the Linear Sampling Method for Three-Dimensional Inverse Electromagnetic Scattering Problems,” *Inverse Problems*, vol. 25, no. 1, p. 015014, 2009.
- [186] Colton, D. and Kress, R., *Inverse Acoustic and Electromagnetic Scattering Theory*, 2d ed., Springer, 1998.

- [187] Devaney, A.J., “Super-Resolution Processing of Multistatic Data Using Time Reversal and Music,” 2000. Available at <http://www.ece.neu.edu/faculty/devaney>.
- [188] Luke, D.R. and Devaney, A.J., “Identifying Scattering Obstacles by the Construction of Non-scattering Waves,” *SIAM Journal on Applied Mathematics*, vol. 68, no. 1, pp. 271–291, 2007.
- [189] Cheney, M., “The Linear Sampling Method and the Music Algorithm,” *Inverse Problems*, vol. 17, no. 4, p. 591, 2001.
- [190] Paige, C.C. and Strakos, Z., “Unifying Least Squares, Data Least Squares, and Total Least Squares,” in *Total Least Squares and Errors-in-Variables Modeling*, Ed. S. v. Huffel and P. Lemmerling, Springer, 2002.
- [191] Wang, Y., Yang, J., Yin, W., and Zhang, Y., “A New Alternating Minimization Algorithm for Total Variation Image Reconstruction,” *SIAM Journal of Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.
- [192] Romberg, J., “Compressive Sensing by Random Convolution,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.
- [193] Ender, J.H., “On Compressive Sensing Applied to Radar,” *Signal Processing*, vol. 90, no. 5, pp. 1402–1414, 2010. Special Section on Statistical Signal & Array Processing.

5.9 | PROBLEMS

1. Generate several small matrices in MATLAB and numerically compute their mutual coherence and restricted isometry constants. Notice how rapidly the problem of estimating the RIC becomes intractable as you increase M , N , and s .
2. Assume that \mathbf{A} represents oversampling of a one-dimensional discrete Fourier transform. Find an analytical expression for the mutual coherence of this dictionary. Verify the result in MATLAB. (Hint: Utilize the expression for the Dirichlet kernel.)
3. We noted in the text that $\mathcal{M}(\mathbf{A}) \leq \mathcal{R}_s(\mathbf{A}) \leq (s - 1)\mathcal{M}(\mathbf{A})$. Prove these bounds. Hint: $\|\mathbf{x}\|_1 \leq \sqrt{s} \|\mathbf{x}\|_2$ if \mathbf{x} is s sparse. Also, notice that $\|\mathbf{A}\mathbf{x}\|_2^2 = \mathbf{x}^H \mathbf{A}^H \mathbf{A} \mathbf{x} = \|\mathbf{x}\|_2^2 + \sum_{i \neq j} x_i^* x_j \mathbf{G}_{ij}$, where $\mathbf{G} = \mathbf{A}^H \mathbf{A}$.
4. The condition number of a matrix is given by the ratio of the largest singular value to the smallest singular value. An identity matrix has condition number 1, and the condition number will approach infinity as a matrix becomes closer to being rank deficient. One might be tempted to conclude that a good condition number indicates that a matrix \mathbf{A} has a good RIP constant. Provide a counterexample to this claim. Verify your counterexample in MATLAB.
5. We have alluded to the idea that mutual coherence is linked to the radar ambiguity function. In particular, the values encoded in the Gramian matrix $\mathbf{A}^H \mathbf{A}$ are samples of the radar ambiguity function. Demonstrate that this relationship holds.
6. Implement the following algorithms in MATLAB or your language of choice. Test the algorithm using a variety of problem sizes. Try generating the \mathbf{A} matrix from a Gaussian distribution, Rademacher distribution (± 1), and as a subset of a DFT matrix.
 - (a) OMP
 - (b) CoSaMP
 - (c) FISTA
 - (d) IHT

7. Generate a phase transition plot comparing the performance of BP, OMP, CoSaMP, and IHT for the noise free case. Select $N = 256$ for your simulations. Use Gaussian random entries for your \mathbf{A} matrix. Be sure to normalize the columns of the \mathbf{A} matrix and to generate a new random \mathbf{A} matrix for each realization. For each algorithm, plot the level curve representing the 50% probability of successful reconstruction. *Hint: You will want to choose a reasonable numerical tolerance and count a realization as a success when the solution is correct to that tolerance. For BP, that is BP_σ with $\sigma = 0$, you can use any of several available solvers found online. SPGL1 can be used, for example.*
8. Obtain the MATLAB implementation of SPGL1 online. Generate a Gaussian \mathbf{A} matrix and a sparse random vector. Attempt to reconstruct \mathbf{x}^{true} from the data \mathbf{y} for a variety of λ values. For each value of λ , solve the problem using SPGL1 with $\sigma = \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\|_2$ and verify that you obtain the same solution. Plot the Pareto curve represented by these results.
9. Suppose that we wish to solve QP_λ with a non-singular weighting matrix \mathbf{W} in place of the scalar weight λ , that is,

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{W}\mathbf{x}\|_1 + \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad (5.61)$$

Demonstrate that we can use any algorithm which is capable of solving QP_λ to obtain the desired solution by solving a modified problem and then applying a simple transformation to the output.

PART II

Synthetic Aperture Radar

CHAPTER 6 Spotlight Synthetic Aperture Radar

CHAPTER 7 Stripmap SAR

**CHAPTER 8 Interferometric SAR and Coherent
Exploitation**

Spotlight Synthetic Aperture Radar

Daniel A. Cook

Chapter Outline

6.1	Introduction	211
6.2	Mathematical Background	214
6.3	Spotlight SAR Nomenclature	220
6.4	Sampling Requirements and Resolution	225
6.5	Image Reconstruction	234
6.6	Image Metrics	240
6.7	Phase Error Effects	244
6.8	Autofocus	250
6.9	Summary and Further Reading	253
6.10	References	255
6.11	Problems	257

6.1 | INTRODUCTION

The decades of development of synthetic aperture radar (SAR) have resulted in a family of remarkable signal processing techniques that are capable of producing imagery whose cross-range resolution is independent of range and much finer than is possible to achieve with any practically-deployable real-beam antenna. SAR was initially conceived in the 1950s by Carl Wiley in the context of stripmap collection. He was, by all accounts, a talented and colorful person [1] whose contributions include the first serious consideration of the feasibility of the solar sail [2]. This early work ignited a legacy of research, development, and practical application that remains strong to this day.

Years later it was realized that under certain conditions the collected radar data can be thought of as sampling the Fourier transform of the scene reflectivity. This thinking produced the spotlight mode, which is the topic of this chapter. Spotlight mode is arguably the dominant type of SAR in current usage. While the stripmap mode is useful for imaging large areas, its resolution is typically more coarse. For this reason, spotlight SAR is commonly employed in military applications where the goal is not to survey a large region but to produce the best possible image of a specific area of interest. This chapter emphasizes spotlight SAR. Yet, in the grand scheme it is important to realize that the various SAR modes—spotlight, stripmap, and inverse—are somewhat artificial distinctions based on categorizing algorithms according to criteria such as application, collection geometry, and

simplifying mathematical assumptions. The principle underlying all synthetic aperture imaging is that of using measurements of a reflected wave field to infer the structure of the distribution of scatterers that created it.

To gain an intuitive understanding of what SAR processing does, imagine a stone dropped into a still pond at an unknown location. If we could measure the waves all around the pond as they reach its edge, it seems reasonable to expect that we could determine the location and amplitude of the initial disturbance given that we know how fast water waves propagate. Not surprisingly, if we measure the waves only at a portion of the pond's edge then we diminish our ability to infer the initial disturbance. The connection to SAR is the principle that cross-range resolution is proportional to the angular interval over which the target is observed. It then becomes simple to compare the resolution achievable by spotlight, stripmap, and inverse SAR collections, as all of these modes function by measuring radar reflections through some angular interval.

SAR belongs to a rather distinguished family of imaging techniques that has garnered several Nobel Prizes. For example, Martin Ryle and Antony Hewish won the 1974 prize for Physics for applying aperture synthesis to radio astronomy. Also, the 1979 prize for Medicine was awarded to Allan Cormack and Godfrey Hounsfield for their work in computer-assisted tomography. We see that SAR imaging techniques are in good company, and it is the author's opinion that their beauty makes them well worth learning.

6.1.1 Organization

In the interest of pragmatism, we narrowly focus on Fourier-domain image reconstruction based on the assumption that spherical radio waves incident upon the scene can be approximated as being locally planar. This approach leads to the polar format algorithm (PFA), which is the most common method of creating spotlight SAR imagery. It is also a good framework for understanding advanced applications such as interferometry and coherent change detection. These are covered in Chapter 8. The objective of this chapter is to discuss spotlight SAR imaging as it is usually encountered in practice. The presentation is succinct, as the extant literature provides a number of superb discussions for readers interested in more details regarding specific topics (e.g., [3,4] and [5]).

We begin with a brief discussion of the Fourier transform and its properties followed by a description of the SAR data collection geometry. The radar sampling constraints in range and cross-range are then derived, along with the achievable image resolution. The discussion then turns to image reconstruction using the polar format algorithm and the tools used for evaluating image quality. The chapter concludes with a look at the image degradation caused by common phase error effects and two common autofocus algorithms used to correct these effects.

6.1.2 Key Points

The key points discussed in this chapter include the following:

- A great deal of SAR imaging can be understood in terms of basic signal processing ideas, such as the Fourier transform and the Shannon–Nyquist sampling theorem.
- SAR collections are subject to sampling constraints dictated both by the signal content and the imaging geometry.
- PFA is the most common spotlight SAR reconstruction algorithm, applied when the radiation incident on the scene can be modeled using planar wavefronts. Other approaches are needed for near-field imaging, such as backprojection or the ω - k algorithm.

- Image quality is predicted using deterministic quantities, such as resolution, as well as statistical quantities, like the contrast ratio, that depend on the average scene content.
- The quality of a particular image can be judged using standardized rating scales.
- Image quality can be degraded by phase errors occurring over the course of the collection. These phase errors can affect the resolution and contrast of an image. They can also introduce false copies of the scene content.
- Measuring the platform motion and compensating for it in the image reconstruction processing is the best way to reduce phase errors. Autofocus algorithms are then applied to the imagery to remove residual defocus.

6.1.3 Notation

The notation used in this chapter is summarized in this section. In certain cases a symbol is used several times with a subscript added to impart specific meaning evident from context. For example, τ_c is used to indicate the length of the transmitted radar signal, while τ_g is used to represent the time gating of the receiver.

Symbols used in this chapter

$\delta\phi$	Angular sampling period (rad)
$\delta k_x, \delta k_y$	Wavenumber sampling period (rad/m)
δr	Range resolution (m)
δx	Cross-range resolution (m)
θ	Angle (rad)
λ	Wavelength (m)
σ	Backscatter coefficient (dimensionless)
τ	Time duration (s)
ϕ	Angle (rad)
ω	Radian frequency (rad/s)
B	Bandwidth (Hz)
c	Speed of light (m/s)
D	SAR image diameter (m)
f	Frequency (Hz)
f_D	Doppler frequency (Hz)
G_{TX}, G_{RX}	Transmit and receive gain (dimensionless)
H	Radar altitude (m)
K_c	Chirp rate (Hz/s)
k, \mathbf{k}	Wavenumber, wavenumber vector (rad/m)
L	Distance (m)
R, r	Range (m)
T, t	Time (s)
v	Aircraft speed (m/s)
x, \mathbf{x}	1-D position, 3-D position (m)

6.1.4 Acronyms

Acronyms used are collected in this section. Aside from standard acronyms such as SNR and FFT, this chapter introduces many terms commonly used in the SAR literature.

Acronyms used in this chapter

AASR	Along track ambiguity to signal ratio
ADC	Analog to digital converter
ASR	Ambiguity to signal ratio
CNR	Clutter to noise ratio
CR	Contrast ratio
DTCR	Distributed target contrast ratio
FFT	Fast Fourier transform
GPS	Global positioning system
GSD	Ground sampled distance
INS	Inertial navigation system
ISLR	Integrated side lobe ratio
LFM	Linear frequency modulation
MLE	Main lobe energy
NESZ	Noise equivalent sigma zero
NIIRS	National image interpretability rating scale
MNR	Multiplicative noise ratio
PDA	Phase difference algorithm
PGA	Phase gradient algorithm
PFA	Polar format algorithm
PRI	Pulse repetition interval
PRF	Pulse repetition frequency
PSR	Point scatterer response
QNR	Quantization noise ratio
QPE	Quadratic phase error
RASR	Range ambiguity to signal ratio
RF	Radio frequency
RVP	Residual video phase
SAR	Synthetic aperture radar
SLE	Side lobe energy
SNR	Signal to noise ratio

6.2 | MATHEMATICAL BACKGROUND

Anyone with an understanding of basic signal processing ideas [6] is well-equipped to achieve a deep understanding of SAR imaging [7,8]. This section introduces the most essential mathematical and signal processing theory needed to understand spotlight mode SAR.

6.2.1 The Fourier Transform

The fundamental mathematical tool required for this chapter is the Fourier transform [9] in one, two, and three dimensions. The principles in multiple dimensions are essentially the same as for a single dimension, so we begin there. The one-dimensional Fourier transform and its inverse are defined as:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt \quad (6.1)$$

and

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \quad (6.2)$$

For a given frequency ω , the complex scalar value $F(\omega)$ of the forward transform (6.1) can be thought of as the inner product of the function $f(t)$ and a complex sinusoid of frequency ω , represented by $e^{j\omega t}$. The negative sign in the exponential of (6.1) is present because of the definition of the inner product for complex-valued functions [10,11]. The collection of values $F(\omega)$ for all ω is called the spectrum of the function $f(t)$. The time and frequency notation used in this section are not the only possible Fourier domains. We will encounter others, and t and ω can be thought of as dummy variables whose purpose is to distinguish the two domains bridged by the Fourier transform.

Any two complex sinusoids are orthogonal (their inner product is zero) as long as their frequencies are not identical. This property is used in the analysis (decomposition) and synthesis (creation) of signals using complex sinusoids as the elemental building blocks. The spectrum $F(\omega)$ represents the amount of each $e^{j\omega t}$ present in $f(t)$. The inverse transform is the reconstitution of the original function $f(t)$ by summing over all the complex sinusoids properly weighted and phased by $F(\omega)$. Understanding the Fourier transform and its basic properties is absolutely essential to SAR processing. An excellent reference is the classic book by Bracewell [9].

We next use the definition of the Fourier transform to derive one of its key properties. We will discover what happens to the transform of the function $f(t)$ when it is shifted by t_0 . First substitute $f(t - t_0)$ into (6.1) and then change the variable of integration from t to $t + t_0$:

$$\begin{aligned} \mathcal{F}\{f(t - t_0)\} &= \int_{-\infty}^{\infty} f(t - t_0) e^{-j\omega t} dt \\ &= \int_{-\infty}^{\infty} f(t) e^{-j\omega(t+t_0)} dt \\ &= e^{-j\omega t_0} \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \\ &= e^{-j\omega t_0} F(\omega) \end{aligned} \quad (6.3)$$

Thus, the Fourier transform of a shifted function is the transform of the unshifted function multiplied by a complex exponential with a linear phase proportional to the amount of the shift. This and other important facts relevant to SAR are collected in Table 6-1.

6.2.2 The Sinc Function

As an introductory example that we'll use extensively later, we find the Fourier transform of the rectangle function defined as:

$$\text{rect}\left(\frac{t}{T}\right) = \begin{cases} 1 & \text{if } |t| \leq \frac{T}{2} \\ 0 & \text{if } |t| > \frac{T}{2} \end{cases} \quad (6.4)$$

TABLE 6-1 ■ Properties of the Fourier Transform Commonly Used in SAR Imaging. The double arrow symbol \Leftrightarrow is used to indicate that two functions are related via the Fourier transform.

Property	Description	Application to SAR
Linearity	$af(t) + bg(t) \Leftrightarrow aF(\omega) + bG(\omega)$, where a and b are constants	Superposition of scatterers in an image
Time scaling	$f(at) \Leftrightarrow \frac{1}{ a } F\left(\frac{\omega}{a}\right)$	Deramp scaling of range axis
Time shifting	$f(t - t_0) \Leftrightarrow F(\omega)e^{-j\omega t_0}$	Displacement of scatterer from scene center
Frequency shifting	$e^{j\omega_0 t} f(t) \Leftrightarrow F(\omega - \omega_0)$	Modulation and basebanding
Area under $F(\omega)$	$f(0) = \int_{-\infty}^{+\infty} F(\omega) d\omega$	Peak value of point scatterer response
Multiplication/convolution	$f(t) * g(t) \Leftrightarrow F(\omega)G(\omega)$	Spectral windowing; phase error effect on point scatterer response
Correlation	$f(t) \star g(t) \Leftrightarrow F(\omega)G^*(\omega)$	Matched filtering, or pulse compression

In this case, the Fourier transform integral is easy to write, as the rect function simply alters the limits of integration of (6.1), where $f(t) = 1$:

$$F(\omega) = \int_{-T/2}^{T/2} e^{-j\omega t} dt \quad (6.5)$$

Next, we carry out the integral and evaluate it at the limits:

$$\begin{aligned}
 \int_{-T/2}^{T/2} e^{-j\omega t} dt &= \left. \frac{e^{-j\omega t}}{-j\omega} \right|_{-T/2}^{T/2} \\
 &= \frac{\cos\left(\omega \frac{T}{2}\right) - j \sin\left(\omega \frac{T}{2}\right) - \cos\left(\omega \frac{T}{2}\right) - j \sin\left(\omega \frac{T}{2}\right)}{-j\omega} \\
 &= T \frac{\sin\left(\frac{\omega}{2} T\right)}{\frac{\omega}{2} T} \\
 &= T \operatorname{sinc}\left(\frac{T}{2\pi} \omega\right) \quad (6.6)
 \end{aligned}$$

The last line of this expression is the sinc function, which is defined as $\operatorname{sinc}(x) = \sin(\pi x)/\pi x$. The word *sinc* is a contraction of the Latin term *sinus cardinalis*, or cardinal sine. The derivation above begins with a rect function in the time domain and obtains a sinc function in the frequency domain. These domains are commonly reversed in radar engineering. The rect function is often used to represent the constant-amplitude spectrum of a transmitted signal. After pulse compression (see the correlation property in Table 6-1 as well as Chapter 2), the spectrum is real-valued. Thus, the signal in the time domain is the sinc function determined by the relationship:

$$\operatorname{rect}\left(\frac{\omega}{2\pi B}\right) \Leftrightarrow 2\pi B \operatorname{sinc}(Bt) \quad (6.7)$$

where B is the bandwidth of the transmitted signal (in Hz, not rad/s).

Figure 6-1 plots the sinc function for several values of B , and we notice two things: First, the maximum value of the sinc function is equal to $2\pi B$ as implied by (6.7). This

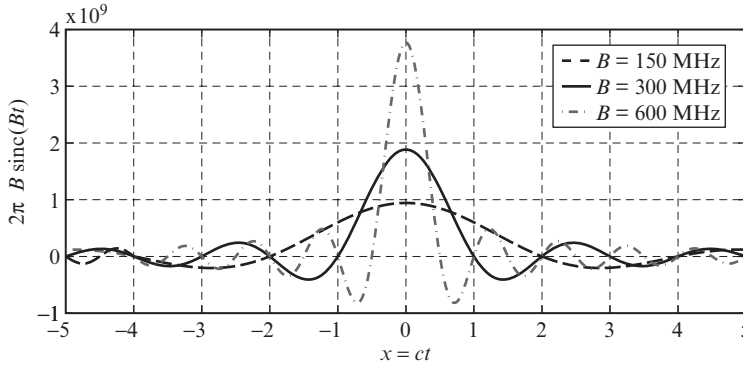


FIGURE 6-1 ■ Sinc function evaluated for several values of the width B of the rect function. Wider rect functions transform into narrower sinc functions, demonstrating that increasing the transmitted signal bandwidth improves its resolution.

also follows from the property in Table 6-1 that states that the area under a function in one domain is given by evaluating its transform at zero. Secondly, we see that wider rect functions (larger values of B) correspond to narrower sinc functions. It turns out that the width of the main lobe of the sinc function is inversely proportional to B , as we now show.

Generally speaking we care about the squared voltage, or power, of electric signals and we ask specifically, “How wide is the sinc function when its power has dropped by half?” We can get an approximate answer using a truncated Taylor series expansion of $\text{sinc}(Bt)$, setting it equal to the appropriate value, and solving for t . We want to solve for $\text{sinc}^2(Bt) = 1/2$, but clearly this is equivalent to solving for $\text{sinc}(Bt) = 1/\sqrt{2}$, which simplifies the math. The first-order approximation to the sinc function is:

$$\begin{aligned} \text{sinc}(Bt) &= \frac{\sin(\pi Bt)}{\pi Bt} \\ &\approx \frac{\sin(\pi Bt_0)}{\pi Bt_0} + \frac{\pi Bt_0 \cos(\pi Bt_0) - \sin(\pi Bt_0)}{\pi Bt_0^2}(t - t_0) \end{aligned} \quad (6.8)$$

Now that we have the Taylor expansion, we need to choose a suitable expansion point t_0 . The normalized sinc function attains its maximum at zero and first crosses zero at $t = 1/B$. In the absence of better information, a logical course of action is to perform the expansion halfway between these points at $t_0 = 1/2B$. Doing so gives:

$$\text{sinc}(Bt) = \frac{2}{\pi} - \frac{4B}{\pi}(t - 1/2B) \quad (6.9)$$

Setting this result equal to $1/\sqrt{2}$ and solving for t we find:

$$t \approx 0.445 \frac{1}{B} \quad (6.10)$$

We learn two important things from (6.10). As expected, we see that the main lobe width is inversely proportional to B so that expanding the rect function causes the width of the corresponding sinc function to get smaller by the same amount. Second, the half-power width is approximately equal to $0.89/B$. Since 0.89 is close to 1, it is common engineering practice to suppress the constant of proportionality and just say that the main lobe width is $1/B$. The rect function is commonly used to represent the magnitude profile of an ideal spectrum. However, other functions may be used for spectral shaping to produce

specific main lobe and sidelobe behavior. Interesting features of numerous such windowing functions are described by Harris [12].

When we let B grow to be arbitrarily large, the rect function becomes infinitely wide and the corresponding sinc function grows infinitesimally narrow. Without rigorous justification of the assertion, we simply state the following transform pair:

$$1 \Leftrightarrow \delta(t) \quad (6.11)$$

where $\delta(t)$ is the well-known Dirac delta function that takes the value 1 at $t = 0$ and zero everywhere else [9]. For our purposes, the important feature of the delta function is that equations (6.3) and (6.11) can be combined to find that its Fourier transform is a complex sinusoid:

$$e^{-j\omega t_0} \Leftrightarrow \delta(t - t_0) \quad (6.12)$$

This property holds in higher dimensions, and it is critical to understanding many aspects of SAR imaging. Each point in a scene can be idealized as a shifted and weighted delta function where the weight corresponds to the reflectivity at the location implied by the shift.

6.2.3 Spatial Frequency and Plane Waves

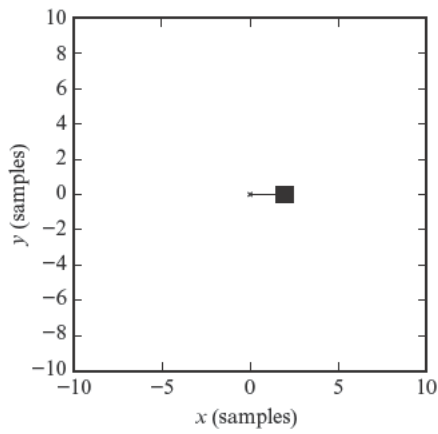
The familiar symbols t and ω are used to represent time and temporal frequency, respectively. Equally useful is the notion of spatial frequency, symbolized as k with units of radians per meter. While temporal frequency ω represents the number of radians a wave cycles through per unit time, the spatial frequency represents the number of radians a harmonic (or single frequency) wave passes through per unit length. The wavenumber is often written as $k = 2\pi/\lambda$, or equivalently $k = \omega/c$ since $c = \lambda f$. The wavenumber vector $\mathbf{k} = [k_x \ k_y \ k_z]^T$ is used to describe the spatial frequency in multiple dimensions.

The Fourier transform of a shifted delta function, given in equation (6.12), also holds in higher dimensions:

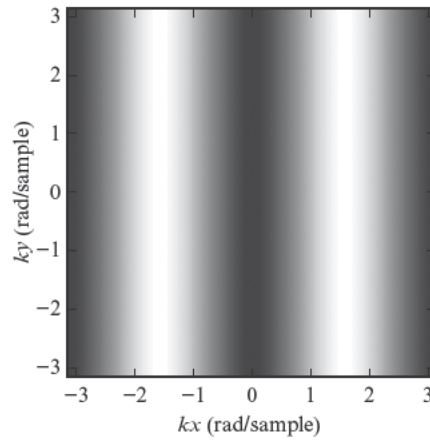
$$\delta(\mathbf{x} - \mathbf{x}_0) \Leftrightarrow \exp\{-j\mathbf{k} \cdot \mathbf{x}_0\} \quad (6.13)$$

The two-dimensional case is obtained when \mathbf{x}_0 has no z component. These two- and three-dimensional complex sinusoids are called plane waves because contours of constant phase are parallel. Some examples are shown in the right-hand column of Figure 6-2. Given that a delta function in one domain transforms into a plane wave in the other, the frequency of this plane wave is proportional to the impulse's distance from the origin. The orientation of the plane wave corresponds to the location of the delta function around the origin. In this fashion each point in the SAR image corresponds to a unique spatial frequency, or two-dimensional complex sinusoid in the Fourier domain.

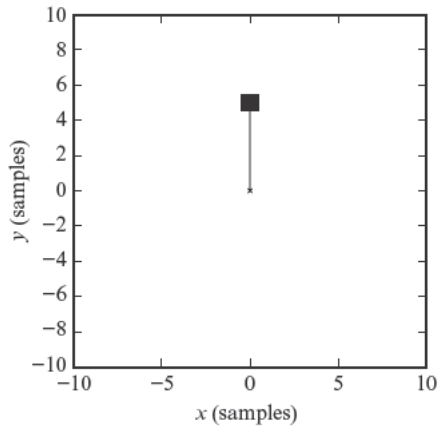
The SAR image is just like any other signal in that it is composed of (and can be expressed as) the weighted sum of sinusoids. SAR practitioners rely heavily on this Fourier-oriented mindset. The relevance to SAR is that we can infer $\delta(\mathbf{x} - \mathbf{x}_0)$ by making measurements in the frequency domain and then Fourier transforming the result to obtain the image. The amplitude and location of the delta function will be shown to correspond to the amplitude of the scene reflectivity at a particular location on the ground. Since the imaging process is linear, superposition holds and we are able to reconstruct the



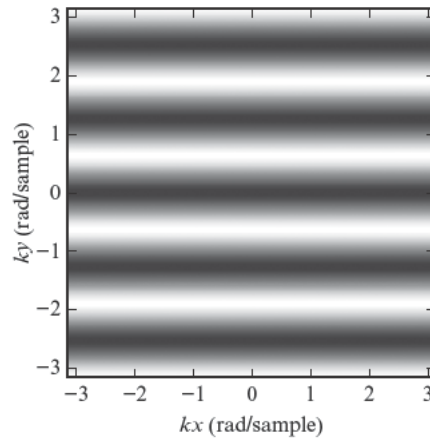
(a)



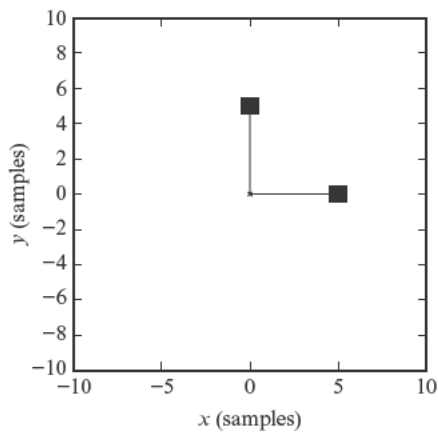
(b)



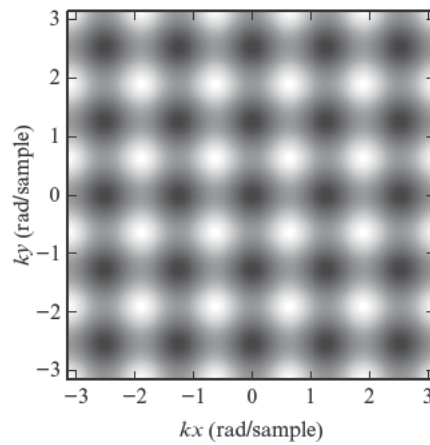
(c)



(d)



(e)



(f)

FIGURE 6-2 ■ The Fourier dual of a complex sinusoid $\exp\{-j\mathbf{k} \cdot \mathbf{x}_0\}$ is a delta function located at \mathbf{x}_0 : $\delta(\mathbf{x} - \mathbf{x}_0)$. The locations of various spatial-domain impulses are shown in the left-hand column, and the right-hand column shows the corresponding two-dimensional Fourier transforms.

entire scene at once. Later we will see that it allows us to easily describe the sampling requirements, resolution, and area coverage of an image.

6.2.4 Fundamental Assumptions of SAR Imaging

Most of SAR processing is based on the following key assumptions:

1. The world can be modeled as a continuum of infinitesimally small point scatterers whose reflections do not affect one another. In other words, we assume that there is no multiple scattering and that the principle of superposition applies to the reflected signals. The linearity of the signals is one reason we can reconstruct imagery without prior knowledge of the scene content.
2. The incident radiation does not penetrate the reflecting surface. Consequently, the scene can be modeled as a sheet which is not necessarily flat, but is infinitesimally thin. It is this model that gives meaning to the interferometric applications of SAR used for deriving topographic information from image pairs. When the scene deviates vertically from a plane, the resulting three-dimensional spectrum has a component whose phase can be used to infer the terrain height. (See Chapter 8 for more information.)
3. The medium is assumed to be infinite in extent, so there are no boundaries to consider. The presence of boundaries can give rise to multipath effects resulting in multiple reflections, or ghosts, in imagery.
4. The scene content does not move during the collection interval of the SAR data. The relative motion between sensor and scene is used to resolve and place targets in cross-range. An object moving against the static background may not appear in the imagery in the correct location. A classic example often cited is the phenomenon of imaging a moving train. If the train's velocity vector has any component along the radar's line of sight, the cars will appear displaced in cross-range. It is possible to get rather disconcerting imagery showing a well-focused train that appears to be riding off its track.

These assumptions are somewhat predictable, representing a commonsense approach to modeling the radar's environment. They also have deeper mathematical meaning, enabling us to obtain simple solutions to the wave equation governing the propagation of electromagnetic waves. In particular, restrictions 1 and 3 give rise to the term e^{-jkR}/R , which is ubiquitous in the radar discipline. It is the Green's function representing the solution to the free space wave equation [13].

Sometimes additional simplifications are made in order to derive efficient algorithms. A common approximation is that the transmitted spherical wavefronts are effectively planar over the scene to be imaged; this assumption is key to the polar format algorithm. Other methods, such as backprojection and the ω - k algorithm recognize the spherical nature of the wavefronts.

6.3 | SPOTLIGHT SAR NOMENCLATURE

Synthetic aperture radars are flexible instruments that are used under a variety of conditions. This section establishes the vocabulary necessary to describe the features of a SAR collection. We will also mention how the collection geometry influences the imagery. Much of this discussion applies to stripmap SAR as well as spotlight.

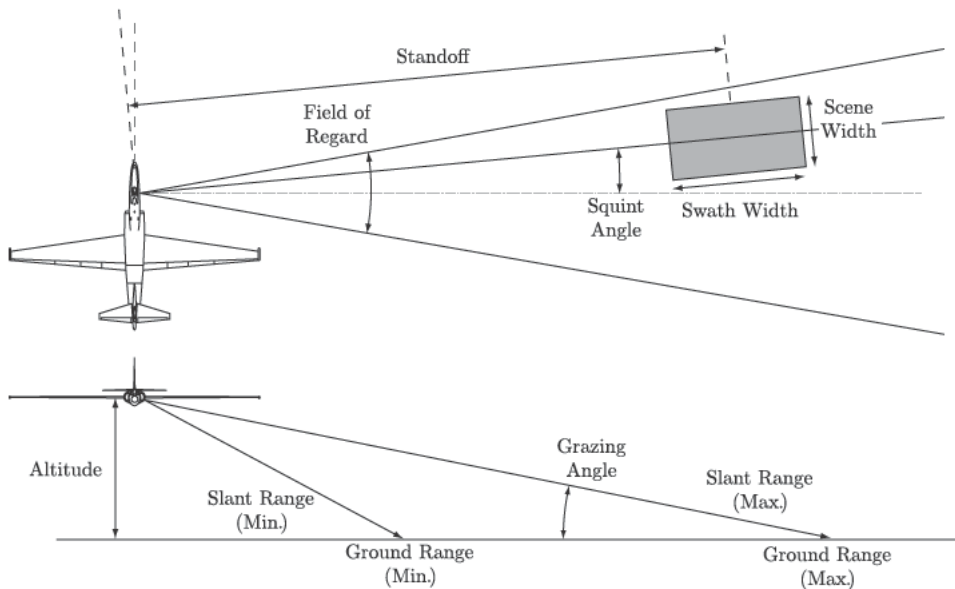


FIGURE 6-3 ■ SAR data collections are generally described using the terms shown here. This drawing is not to scale. The dimensions of the SAR image are typically much smaller than the range to the platform.

An airborne SAR collection geometry is illustrated in Figure 6-3. The distance from the sensor to a point on the ground is known as the slant range. Projecting along wavefronts into the plane of the ground gives the ground range. The range between the aircraft and scene center is often called the standoff distance. The standoff does not impact the achievable cross-range resolution as it does for real-beam imaging, as SAR resolution is independent of range. However, it does contribute to determining the transmit power and antenna gain requirements. It also determines the time required to collect the synthetic aperture data. Longer ranges require more integration time for a given integration angle. Increasing the collection time means that platform position error estimates have more time to accumulate, which can affect image quality.

The angle between the vector pointing from the radar to the aim point and the locally horizontal plane at the aim point is called the grazing angle. Choice of the grazing angle is often determined by the terrain reflectivity and shadowing characteristics as well as the desire to obtain an image that has the appearance of an overhead plan view. The grazing angle is also limited by the maximum altitude achievable by the host aircraft. Typical grazing angles range from 20° to 60° but can be much lower if necessary. Note that the grazing angle changes over the range swath. This effect is safe to ignore when the standoff is much greater than the ground range swath width, as is usually the case.

Very shallow grazing angles result in low terrain backscatter and thus can suffer from low signal-to-noise ratio (SNR). As described in Section 5.2.2 of *Basic Principles* [14], terrain reflectivity is often modeled as being proportional to the sine of the grazing angle. This is known as the constant gamma model. Furthermore, much of the scene may be obstructed since low grazing angles result in very long shadows when the terrain is not flat. On the other hand, overly steep grazing angles result in imagery that can be difficult to interpret, since the image one sees is the result of projecting along wavefronts into the slant plane as depicted in Figures 6-4 and 6-5 and explained in detail in Chapter 8. Figure 6-6 shows a pair of images taken by the German TerraSAR-X satellite while operating in spotlight mode. These show the open-pit copper mine in Chuquicamata, Chile. The optical

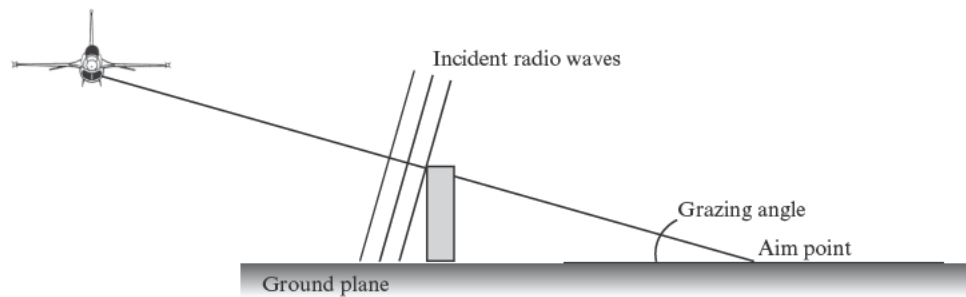


FIGURE 6-4 ■ All of the reflections at the same range are received at the same time. Thus, all reflections seen by a wavefront are projected into the same range bin. Depicted here is the layover effect, in which the top of an elevated terrain feature appears at a closer range than its base.

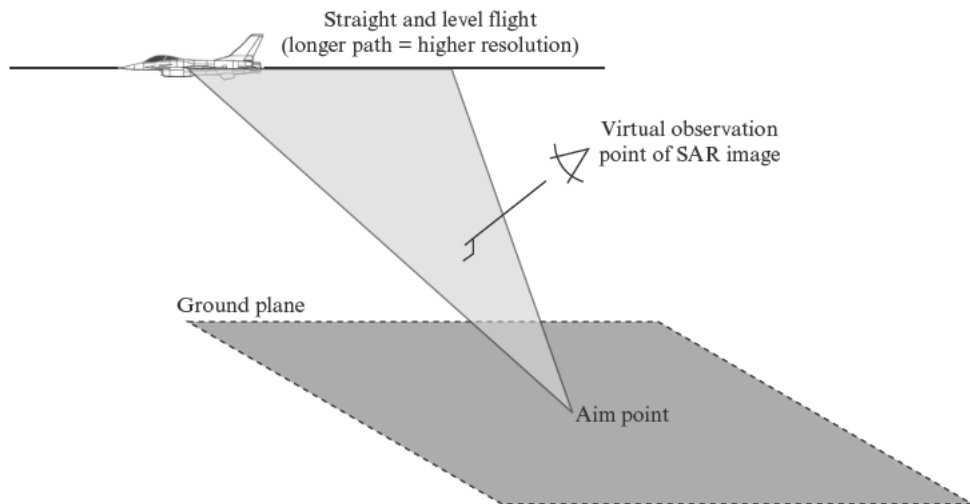


FIGURE 6-5 ■ The image plane of a spotlight SAR is defined by the aim point and the sensor's flight path.

image shows that the sides of the pit are fairly uniform, yet the pit exhibits severe layover in the SAR imagery. Figure 6-7 shows another interesting effect. Here, an airborne Ka-band radar images a pair of aircraft that are largely invisible because the smooth surfaces cause the incident radiation to reflect away from the radar. The backscatter is caused primarily by sharp corners and discontinuities such as rivet lines. In this image, the shadows provide clear outlines of the aircraft.

The radar collects SAR data by flying with the antenna continuously steered to point toward the same location on the ground, usually called the aim point. The collection lasts as long as is required for the platform to fly through the integration angle needed to achieve the desired cross-range resolution. Spotlight SAR data are almost always collected by flying straight and level. However, spotlight-mode processing algorithms are usually derived by assuming that the radar flies in a fixed-radius circle about the aim point. Consequently, each receive window is time delayed to correspond to a circular flight path. Some spotlight radars actually do fly circular trajectories, but this is not the norm [15].

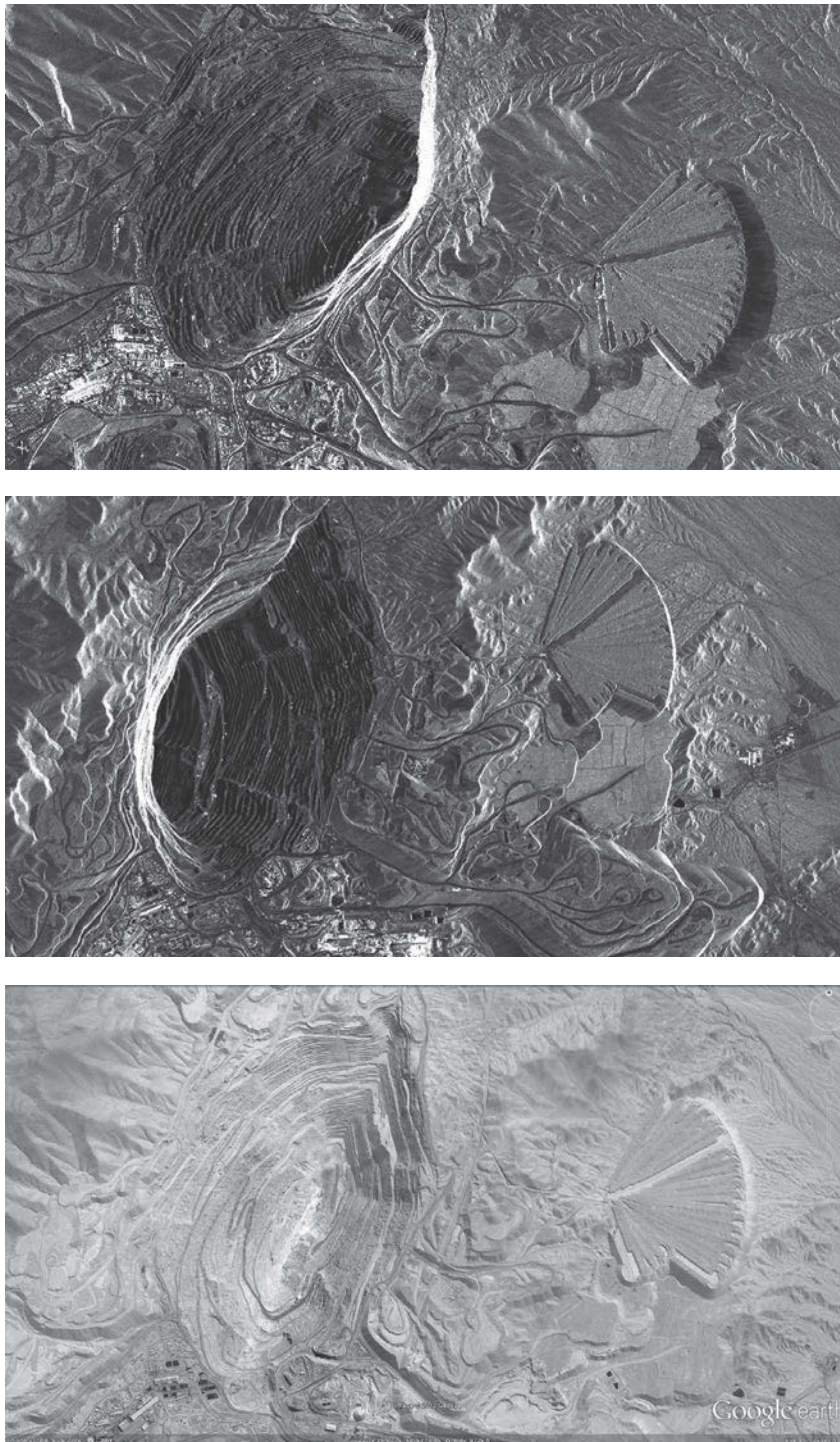


FIGURE 6-6 ■
TerraSAR-X imagery of the Chuquicamata copper mine in Chile (copyright 2009 Infoterra GmbH).
Top: Image taken on ascending pass, with radar west of the scene and 55° grazing angle.
Middle: Image taken on descending pass, with radar east of the scene and 59° grazing angle.
Bottom: Optical image of the mine (copyright 2012 Google and GeoEye).

FIGURE 6-7 ■
Ka-band SAR image
of an airplane and
helicopter on static
display. (Courtesy of
Sandia National
Laboratories)



The squint angle, shown in Figure 6-3, describes the antenna pointing angle relative to broadside. It is usually taken to be the pointing angle at the middle of the collection. When the radar points normal to the flight path, directly port or starboard, this is known as a broadside geometry. We define the squint angle relative to broadside, but it is also common to reference the squint angle to the velocity vector of the radar. Squint angle may be measured in the ground plane. It is also sometimes equated to the Doppler cone angle, or the angle between the radar's velocity vector and the line of sight vector pointing to scene center. The range of available squint angles is known as the field of regard and depends on the radar hardware. Some mechanically gimbaled antennas offer a very wide field of regard. On the other hand, an electronically scanned antenna fixed to the side of an aircraft fuselage would have a narrower field of regard.

The squint angle can impact imagery in several ways. Assuming that the aircraft speed is constant and the ground range is fixed, it can be shown that squinted collections require more time to collect. The necessary aperture time is approximately $T_{\text{squint}} = T_{\text{broad}} / \cos(\theta_{\text{cone}})$, where T_{broad} is the time required to collect the data when looking broadside, and θ_{cone} is the cone angle associated with the look direction. Observe that T_{squint} is always greater than T_{broad} when the radar looks anywhere other than broadside. As was the case with longer ranges, increased aperture time means that the aircraft motion measurement has more time to drift, potentially making motion compensation and autofocus more difficult. Another effect of squinted collections is that vertical objects will always appear to lay over toward the closest point of approach of the radar. In other words, the layover is always perpendicular to, and points toward, the radar's ground track. Meanwhile, the shadows always fall along the radar's line of sight. Thus, layover and shadow are antiparallel only for broadside collections. This effect is described in detail in Appendix C of [5].

Moving to the scene itself, the swath depth or swath width is the downrange extent of the image. The term *swath* probably originated with stripmap imaging, where the radar creates a swath of imagery as it passes over the earth. Given the speed of light, the swath depth determines when the signal recording starts and ends. This timing must account for the projection of the ground into the slant plane (see Figure 6-4). If the swath depth on the

TABLE 6-2 ■ Typical Ranges of Key Spotlight SAR Collection Parameters

Frequency	High: X, Ku, and Ka bands (fine resolution) Low: UHF, VHF, and L bands (coarse resolution, FOPEN)
Grazing angle	$20^\circ - 60^\circ$
Squint angle	$\pm 45^\circ$, (0° is broadside)
Integration time	<1 second to 1 minute
Resolution	10 m – 0.1 m
Scene size	100 m – 100 km
Standoff	1 km – 100 km

ground is L , then the radar record window must be at least $T_{rec} = 2L \cos(\theta_{grz})/c$. To this time we must also add the length of the transmitted pulse to ensure that the farthest point in the image can be properly pulse compressed.

The swath depth, shown in Figure 6-3, is bounded by the footprint on the ground of the radar's vertical beamwidth. For maximum power efficiency, the radar would illuminate only the scene of interest. Generally speaking, however, the downrange footprint is larger than the imaged scene. Similar considerations apply in the cross-range dimension where the radar's horizontal beamwidth limits the cross-range extent, or width, of the scene.

Some example parameters for spotlight synthetic aperture radars are given in Table 6-2. These are not intended to define the performance envelope of spotlight SAR. Rather, their purpose is to give a sense of typical operational characteristics. For example, grazing angles much lower than 20° may be used when the terrain allows. Furthermore, it is important to point out that the low-frequencies and wide-beams characteristic of foliage penetrating radar often preclude the use of PFA for image reconstruction because wavefront curvature across the scene cannot be ignored. These collections require the use of algorithms, such as backprojection [16], that are more accurate but much more computationally expensive.

6.4 | SAMPLING REQUIREMENTS AND RESOLUTION

SAR sampling and resolution concepts are similar, regardless of the collection mode. The resolution achievable by the radar is inversely proportional to the angle through which the scene is observed, commonly called the integration angle. Because SAR collections are made from moving platforms, it is also common to refer to the integration time of the synthetic aperture. It is easy to get confused on this point, but remember that the data could be collected over an arbitrarily long time provided the scene content is static. In principle, a helicopter stopping to hover at each pulse location would give the same image as a trans-sonic fixed-wing aircraft that collects the same set of pulses. SAR imaging exploits spatial configuration changes from pulse to pulse, regardless of when they are collected. SAR data collection might be thought of as a geartrain: all that matters is that the teeth mesh properly. Once this happens the machine works the same, whether it turns slowly or quickly. In this section we will establish the basic machinery of SAR

by examining the sampling requirements and the factors that affect the image size and resolution. These relationships are explained using standard one-dimensional signal processing concepts because the range and cross-range properties of SAR are essentially separable.

6.4.1 Temporal Sampling Requirements and Range Resolution

Temporal (or fast-time) sampling and resolution works the same for SAR as it does for any other radar. See, for example, Chapter 2 of this book as well as Chapter 20 of *Basic Principles*. Myriad waveforms are used within the radar discipline, but one is the canonical SAR waveform: the linearly swept frequency-modulated chirp, or LFM for short. The LFM is detailed in Chapter 2, but we review the important features here. The LFM chirp is swept in frequency from f_{min} to f_{max} , meaning that its bandwidth is $B_c = f_{max} - f_{min}$. The amplitude of the pulse is kept constant throughout its sweep. As we see from the Fourier transform relation (6.7) between the rect and sinc functions, increasing the bandwidth of the transmitted pulse results in finer range resolution.

The sweep rate is constant and equal to $K_c = B_c/\tau_c$ where τ_c is the pulse duration. The choice of τ_c is often determined by the transmit power capacity of the radar. The LFM pulse compression gain is equal to its time–bandwidth product $B_c\tau_c$, so longer transmit pulses yield better SNR. The longer pulses imply a higher duty factor $D = \tau_c/T$ for a fixed pulse repetition interval T . For a fixed amplitude, increasing τ_c implies increasing the average power output of the radar. We will see later that the imaging geometry also places constraints on the coupling between the pulse repetition frequency (PRF) and τ_c . It is the radar designer’s role to find a waveform that satisfies these conditions while achieving the desired image quality.

6.4.1.1 Deramp on Receive

The RF bandwidth necessary to achieve fine range resolution has historically been beyond the ability of analog-to-digital converters (ADCs). For example, 0.3 m range resolution implies an RF bandwidth of at least 500 MHz, since the range resolution is given by $\delta x_{3dB} = c/2B_c$. While technology is always improving, ADCs capable of high-bandwidth digitization with sufficient bit depth may be difficult to obtain or prohibitively expensive. The concept of dechirp-on-receive, more commonly called deramp or stretch processing, provides a workaround if the transmitted signal is an LFM chirp with a large time–bandwidth product [3]. Deramp reception permits the use of an ADC rate that is significantly lower than that implied by the transmitted bandwidth and the ultimate range resolution in the slant plane. A deramp receiver multiplies (or mixes) the incoming reflection with a delayed version of the transmitted chirp, and this operation effectively pulse compresses and takes the Fourier transform of the received signal. For a fixed ADC rate, sampling density in the frequency domain corresponds to signal length in the time domain. Thus, the sampling requirements of the deramp receiver are determined by the swath depth, not the signal bandwidth.

The copy of the transmitted LFM used for mixing in the receiver has a delay corresponding to some reference range r_0 , usually the scene center. After mixing, the reflection from a particular range r becomes a tone whose frequency is proportional to the distance from scene center and whose amplitude and phase are determined by the complex-valued reflectivity $f(r)$ at that point on the ground. The sampling rate of the ADC needs only to

support the bandwidth implied by the minimum and maximum frequencies corresponding to the ends of the desired range swath. We present the basic ideas behind deramp reception in this section. More detail is found in Chapter 2 and specifically as applied to synthetic aperture imaging in [3,5,17].

The transmitted LFM chirp signal is

$$p(t) = \text{rect}\left(\frac{t}{\tau_c}\right) \exp\{-j(\omega_0 t + \pi K_c t^2)\} \quad (6.14)$$

where τ_c is the transmitted chirp duration (s), B_c is the chirp bandwidth (Hz), and $K_c = B_c/\tau_c$ is the sweep rate (Hz/s). The reflected signal is described by convolving the transmitted pulse with the scene reflectivity, but we will consider only the return from a single range r :

$$s_0(t) = f(r) \cdot p\left(t - \frac{2r}{c}\right) \quad (6.15)$$

where $f(r)$ is the scene reflectivity associated with range r . A deramp receiver mixes $s_0(t)$ with a delayed chirp having the same sweep rate as the transmitted chirp, but with a longer duration τ_g , which we term the range gate. The result of mixing is

$$\begin{aligned} s(t) &= f(r) \cdot p\left(t - \frac{2r}{c}\right) \cdot p^*\left(t - \frac{2r_0}{c}\right) \\ &= \text{rect}\left(\frac{t - 2r_0/c}{\tau_c}\right) \cdot f(r) \cdot \exp\left\{-j\left(\omega_0\left(t - \frac{2r}{c}\right) + \pi K_c\left(t - \frac{2r}{c}\right)^2\right)\right\} \\ &\quad \cdot \exp\left\{+j\left(\omega_0\left(t - \frac{2r_0}{c}\right) + \pi K_c\left(t - \frac{2r_0}{c}\right)^2\right)\right\} \\ &= \text{rect}\left(\frac{t - 2r_0/c}{\tau_c}\right) \cdot f(r) \\ &\quad \cdot \exp\left\{-j\left(\frac{2}{c}\left[\omega_0 + 2\pi K_c\left(t - \frac{2r_0}{c}\right)\right](r - r_0) - \frac{4\pi K_c}{c^2}(r - r_0)^2\right)\right\} \end{aligned} \quad (6.16)$$

The frequency of the reflected signal located at r after deramp processing is found by taking the time derivative of the phase in (6.16):

$$f_{\text{deramp}} = \frac{\omega_{\text{deramp}}}{2\pi} = \frac{2K_c}{c}(r - r_0) \quad (6.17)$$

We see that each range maps into a unique frequency, or tone, that is proportional to the range between the scatterer of interest and the deramp reference range r_0 . Thus, the reflectivity profile $f(r)$ can be recovered by Fourier transforming $s(t)$. The range resolution of our estimate of $f(r)$ is inversely proportional to the time over which f_{deramp} is observed, with the constant of proportionality being $(c/2K_c)^{-1}$. This fact follows from the time–frequency scaling property of the Fourier transform (see Table 6-1). If the deramp time gate τ_g is sufficiently long, the observation time of f_{deramp} is equal to or greater than the pulse length. Thus, the range resolution is found to be $c/2K_c\tau_c = c/2B_c$.

Next, we wish to find the bandwidth of the signal leaving the deramp stage. Taking the range swath to be $r \in [-R_s/2, +R_s/2]$, we find from (6.17) that the bandwidth is

$$\begin{aligned} B_{\text{deramp}} &= \frac{2K_c R_s}{c} \\ &= \frac{2B_c R_s}{c\tau_c} \\ &= B_c \frac{\tau_s}{\tau_c} \end{aligned} \quad (6.18)$$

where τ_s is the two-way time required for light to traverse the range swath R_s . This result is one of the most important properties of the deramp processor as applied to fine resolution SAR. Assuming that we transmit a pulse whose duration is longer than the time required to traverse the range swath, we find that the bandwidth leaving the deramp operation is smaller than the transmitted bandwidth. Therefore, we see that the ADC sampling rate required by the deramp receiver is much less than the transmitted RF bandwidth provided that $\tau_s/\tau_c \ll 1$.

Until now, we have ignored the quadratic term $4\pi K_c(r - r_0)^2/c^2$ in the last line of (6.16). This term is known as residual video phase (RVP), and it can be thought of as the difference between the deramp signal and the actual Fourier transform of the reflected radar return. Considering only a single pulse and the reflection from a single range r , we see that the RVP does not depend on time and is a complex-valued constant that can be safely ignored when using a deramp receiver. However, SAR is a two-dimensional operation that exploits the changing range between sensor and target to achieve cross-range resolution. From this viewpoint we find that the RVP changes from pulse to pulse as the sensor flies by a given location on the ground. The result is a quadratic phase error that can corrupt the imagery, as discussed in Section 6.7. Fortunately this error is deterministic, and its compensation, known as range deskew, is addressed in [4].

Let us now examine the effect of the deramp operation on the range resolution, δr_{3dB} . As always, it is given by the speed of light divided by twice the bandwidth:

$$\begin{aligned} \delta r_{\text{3dB}} &= \frac{c}{2B_c} \\ &\leq \frac{c}{2K_c\tau_g} \end{aligned} \quad (6.19)$$

Of course, if $\tau_g > \tau_c$ we cannot realize any additional bandwidth beyond what is given by $K_c\tau_c$. Equation (6.19) tells us that the time gating employed by the deramp receiver can, at best, achieve the resolution implied by transmitted bandwidth B_c . If $\tau_g \geq \tau_c + \tau_s$, then the deramp accounts for the delay across the scene and can process the full LFM reflected by the near and far edges (sometimes called the heel and toe) of the swath. This ensures that the best possible resolution is achieved at all ranges [4]. In situations where $\tau_s \ll \tau_p$ the delay across the scene is small and the effect on the range resolution will be negligible if we happen to set $\tau_g = \tau_p$ [5].

The collection of range profiles over a SAR collection is the raw data used for processing. It is sometimes referred to as phase history data or video phase history data. These terms are dated, but they persist and may be encountered in practice. Modern data acquisition systems capture both the magnitude and phase of the radar reflections. In the early days of radar the data were recorded onto film, which required the signals to be

mixed down to frequencies that could be displayed by a cathode ray tube to which the film was exposed. Phase history data are also sometimes called range history data, since phase is proportional to range and because plotting this information shows the range to a scatterer as a function of time.

One must generally take care with terminology to avoid confusing the domains in which the data reside. For example, the raw data coming from a deramp receiver can be thought of as existing over RF frequency (or fast time frequency) and pulse number (slow time). If the receiver directly samples the incoming signal, then the data exist in range (fast time) and pulse number (slow time). The two domains are equivalent, being separated only by a Fourier transform in the range dimension that can be realized digitally or through analog deramp reception.

Since deramp reception is so common, spotlight SAR data are often assumed to be collected in the Fourier domain and already matched filtered. We will use this fact later in our discussion of image reconstruction via the polar format algorithm. As a final note we point out that it is common parlance to use the phrase pulse compression to refer to the act of inverse Fourier transforming the deramped signal to obtain the range profile. Strictly speaking, however, pulse compression is a synonym for matched filtering. A signal may be pulse compressed regardless of whether it is in the time or frequency domain. The fact that deramp accomplishes both matched filtering and the Fourier transform can make this point confusing.

6.4.2 Along-Track Sampling Requirements

We have already seen that a SAR data collection comprises a set of pulses that are transmitted and received as the sensor flies along its trajectory. It stands to reason that the along-track signal is subject to some sampling rate constraint just as any other signal would be. This means that the Shannon–Nyquist sampling theorem is lurking nearby. In everyday engineering conversation, the sampling theorem states that we need to sample at least twice the rate of the highest frequency present in the signal. The application to SAR spatial sampling becomes clearer if we say this another way: we need to sample at least twice per wavelength. The spatial sampling requirement comes from the need to sample at or before the point at which the change in phase (two-way range) to any visible scatterer changes by $\lambda/2$. Since $\lambda = c/f$, the highest frequency has the shortest wavelength, in turn driving the sampling constraint.

Recall the earlier assertion that SAR can be thought of purely in spatial terms. We now apply this statement to the problem of determining the along-track spatial sampling requirement. Consider a simplification of the SAR collection in which the antenna beam footprint on the ground is idealized as being equal to zero outside of the main lobe. The footprint generally has an elliptical shape, but it is drawn as a circle in Figure 6-8. Either because of the flight path or through motion compensation, a spotlight collection is tantamount to flying in a circle about the aim point (i.e., the center of the scene), so we can also think of the sensor as being stationary while the illuminated scene rotates as if it were on a turntable. The question we need to answer then becomes, “How much can the scene rotate before we need to record another pulse?” The answer is that the angle $\delta\phi$ between pulses must be small enough to ensure that the two-way path from the antenna to any point in the scene changes no more than one-half of a wavelength. The sensor effectively flies in a circle centered on the aim point, so the distance to scene center never changes (we can think of the reflection from the aim point as a DC signal). The maximum path

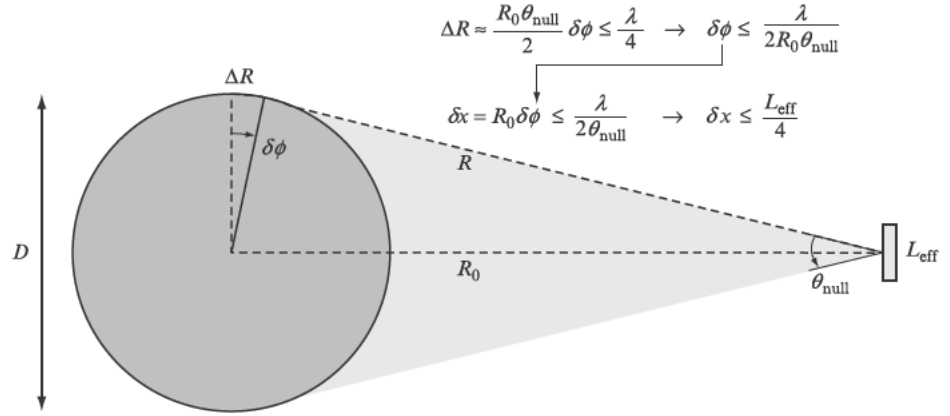


FIGURE 6-8 ■ The cross-range (along-track) sampling requirement for spotlight SAR is given by the requirement that the two-way path to any scatterer in the scene cannot change by more than one-half wavelength between pulses.

length change comes from the edge of the scene, $D/2$ from the center, and located $\pm 90^\circ$ with respect to the radar's line of sight. One of these points is shown in Figure 6-8. We anticipate that $\delta\phi$ is going to be a small angle, so the interpulse path change constraint can be written as

$$\Delta R \approx \delta\phi \frac{D}{2} \leq \frac{\lambda}{4} \quad (6.20)$$

Recognizing that the scene diameter is less than or equal to the range to scene center times the antenna beamwidth, we substitute $D = R_0 \theta_{null}$ to express the limit on ΔR as

$$\begin{aligned} \frac{\delta\phi R_0 \theta_{null}}{2} &\leq \frac{\lambda}{4} \\ \delta\phi &\leq \frac{\lambda}{2 R_0 \theta_{null}} \end{aligned} \quad (6.21)$$

This equation tells us the maximum angular separation between pulses as the radar flies around the scene. Since $\delta\phi$ is small, we can express the angular sampling requirement as a linear, or along-track, sampling requirement by simply computing the distance $\delta x = R_0 \delta\phi$:

$$\delta x = R_0 \delta\phi \leq \frac{\lambda}{2 \theta_{null}} \quad (6.22)$$

We get a different form if we substitute the common expression for the null-to-null beamwidth of an antenna $\theta_{null} \approx 2\lambda/L_{eff}$

$$\delta x \leq \frac{L_{eff}}{4} \quad (6.23)$$

This along-track sampling condition should be considered to be a minimum, as it can cause a considerable amount of sidelobe energy to be aliased into the image. A thorough and accessible discussion of along-track sampling and the resulting spatial aliasing can be found in [17,18]. Chapter 7 shows that this result holds for stripmap collections as well, reinforcing the idea that the distinctions between synthetic aperture imaging modalities are largely a matter of convenience and that they are all connected on a deeper mathematical level.

As promised, we have proven that the along-track sampling requirement for a SAR has nothing to do with time or with the platform speed. It is important to consider time, however, since practically all radar designs specify the pulse repetition interval (PRI) as opposed to the distance traveled between pulses. The maximum PRI is found by dividing δx by the fastest expected platform speed v : $\text{PRI} = \delta x / v$. The PRI has units of seconds, and taking its reciprocal gives the required minimum PRF in hertz.

Many radar engineers like to think in terms of Doppler frequencies, so we can expand the discussion of time dependence even further by recalling that the Doppler frequency shift of a reflected signal is given by $f_D = 2v_{LOS}/\lambda$, where v_{LOS} is the relative sensor-target velocity vector projected onto the line-of-sight unit vector pointing from the sensor to the target. In the context of Figure 6-8 where the beamwidth is a small angle, we find that $v_{LOS} \in [-v \sin(\theta_{null}/2), +v \sin(\theta_{null}/2)] \approx [-v\theta_{null}/2, +v\theta_{null}/2]$ and that the PRF has to be at least equal to twice the highest Doppler frequency present within the illuminated footprint. Since the footprint is symmetric, we might also say that the PRF must be equal to or greater than the Doppler bandwidth of the scene: $\text{PRF} \geq B_D = 2v\theta_{null}/\lambda$.

This requirement for the PRF to be equal to or greater than the Doppler bandwidth of the ground reflections holds even for squinted collections. The reason for this is simple and can be quickly understood through an analogy to ordinary signal processing. The most common statement of the sampling theorem says that we must sample at twice the highest frequency present in the signal. This condition is sufficient to ensure there is never any aliasing of any frequencies present in the signal. However, suppose we know ahead of time that the frequency content is limited to some region $f \in [f_{min}, f_{max}]$ not centered at zero. Then, there are no frequencies lower than f_{min} and the sampling rate needs to be equal to or greater than only the bandwidth, not the highest frequency present. A squinted radar beam acts as a bandpass filter in Doppler (or spatial) frequency.

6.4.3 PRF Constraints

The preceding discussion might lead us to conclude that choosing the SAR PRF is simple enough: just sample the Doppler bandwidth and move on to other problems. Unfortunately, it isn't that easy. Other factors can influence the ultimate choice of PRF. Namely, we must balance the impacts of range and azimuth ambiguities to choose a PRF that yields suitable image quality. Furthermore, there are hard constraints on the PRF imposed by the fact that one cannot transmit a pulse while receiving the echoes from the desired range swath.

First is the fact that the beamwidth isn't equal to zero outside the main lobe. The radar senses all Doppler frequencies $f_D \in [-2v/\lambda, +2v/\lambda]$, although those outside the main lobe are severely attenuated. The antenna pattern behaves just like a low-pass filter for broadside collection or a band-pass filter for squinted collections. We know there is no such thing as a filter with total out-of-band rejection. In fact, many common beam patterns resemble the sinc function, which has terrible rejection properties. As with any other signal, the Doppler frequencies (or along-track spatial frequencies) sampled by the SAR are aliased to some extent, and this aliased energy appears in the image as background noise that reduces contrast [18]. Increasing the PRF reduces the amount of aliased energy that can corrupt the image. However, increasing the PRF implies recording data at a faster rate and transmitting more average power, both of which cost money. Thus, an important part of the design process is to find the minimum PRF that results in the acceptable amount of aliased energy. There is no set answer to this question. A good rule-of-thumb is to set

the PRF at least as high as the Doppler bandwidth implied by the null-to-null width of the main lobe of the antenna, as we did earlier. A more detailed analysis is given in Section 6.6.1.1 where we weigh the impact of azimuth ambiguities in relation to other sources of noise.

Range ambiguities can also arise from the fact that the radar transmits a periodic signal. Therefore, the reflections received from range $R = c\tau/2$ will also contain contributions from ranges $R_n = c\tau/2 + ncPRI/2$, where n is an integer indicating the number of reflections that can occur between the radar and the horizon. These additional unwanted reflections are called range ambiguities since the radar cannot tell which one of a discrete set of ranges caused a particular reflection. For airborne SAR the standoff range is relatively short and the aircraft speed is slow enough such that the ambiguous ranges all occur beyond the imaged scene, and $n \geq 0$. In other words, the pulse being received is typically the last one transmitted. On the other hand, spaceborne SAR is characterized by long standoffs and very fast platforms. For this reason, the PRFs needed to prevent azimuth ambiguities result in the condition where multiple pulses are in flight simultaneously. The pulse being received at any given time was actually transmitted several PRIs earlier. This scheme can work for either spaceborne or airborne SAR as long as the signal is not corrupted by ambiguous ranges between the radar and the image scene. Whether this occurs is a function of the collection geometry and the vertical beam pattern.

In addition to these considerations, there are timing constraints that forbid certain values of the PRF [3,19]. First, the radar cannot transmit while it is receiving signals from the range swath of interest. The interference of transmission upon reception is called eclipsing. Second, it is common practice to ensure that the radar timing gates out the return from directly beneath the sensor. This is known as the nadir return or the ground flash. Although the radar may be pointed well away from vertical incidence, the nadir return is generally assumed to be significant because the ground is an enormous reflector even when located in the antenna sidelobes. If only a single pulse is in the air at a given time, then the nadir flash is automatically gated out since it occurs well before the reflection from the imaged swath. This is usually the case for airborne SAR.

The PRF timing constraints are written in the following set of equations where R_{near} and R_{far} are the slant range from the radar to the near and far edge of the scene, f_p is the PRF, τ_p is the duration of the transmitted pulse, and τ_{rp} is the duration of the receiver protection window [3]. That is, the transmitted pulse is padded by τ_{rp} at the beginning and end to prevent any damage to the receiver that might be caused by leakage from the transmitter. Any valid choice for the pulse repetition frequency must satisfy all of these conditions:

$$\text{frac} \{ 2R_{near}f_p/c \} / f_p > \tau_p + \tau_{rp} \quad (6.24a)$$

$$\text{frac} \{ 2R_{far}f_p/c \} / f_p < \frac{1}{f_p} - \tau_{rp} \quad (6.24b)$$

$$\text{int} \{ 2R_{far}f_p/c \} = \text{int} \{ 2R_{near}f_p/c \} \quad (6.24c)$$

where $\text{frac}\{\cdot\}$ indicates the fractional part of its argument and $\text{int}\{\cdot\}$ indicates the integral part. Equation (6.24a) enforces the constraint that a new pulse cannot be sent if its reflection will overlap a previous reflection from the near edge of the scene. Likewise, equation (6.24b) forces the radar to terminate reception from the far edge of the scene τ_{rp}

seconds prior to transmitting the next pulse. These two requirements are coupled by (6.24c) to require the same PRF for both (6.24a) and (6.24b).

The constraints on the nadir return can also be written as

$$2H/c + i/f_p > 2R_{far}/c \quad (6.25a)$$

$$2H/c + \tau_{nadir} + i/f_p < 2R_{near}/c \quad (6.25b)$$

where H is the altitude of the radar [3]. Equations (6.25a) must be satisfied for any integer value $i = 0, \pm 1, \pm 2, \dots, \pm n_h$. The value n_h is the maximum number of pulses that can occur before the signal passes the radar horizon, and τ_{nadir} is the duration of the reflection from the earth's surface directly beneath the radar. The exact value of τ_{nadir} depends on the terrain, but $\tau_{nadir} = 2\tau_p$ is a common value. Equation (6.25a) requires the nadir return to occur later than the reflection from the far edge of the scene, while (6.25b) forces the nadir return to extinguish before reflection from the near edge of the scene.

Both eclipsing and nadir return timing restrictions can be combined into a graphical representation that applies for a single imaging geometry. That is, the timing possibilities depend on the altitude and the squint and grazing angles. These graphs are sometimes called basket-weave plots or zebra plots. An example is shown in Figure 6-9, which is based on parameters from the Shuttle Imaging Radar B (SIR-B) mission. This diagram shows us that some timings result in multiple pulses in flight between the radar and the ground. Similar plots may be constructed for airborne radars, but they are a bit less interesting since airborne SAR typically has only a single pulse in flight at a time.

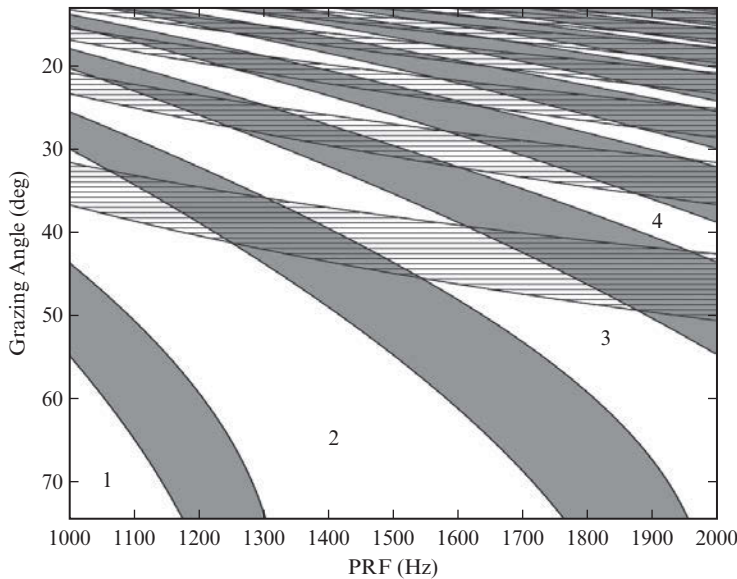


FIGURE 6-9 ■ Accessible PRF values for the SIR-B radar. (Figure adapted from [19] and [3].) The gray regions are PRFs forbidden by eclipsing from the ground range swath, and the hatched areas are denied by the nadir reflection. Usable PRFs are the remaining white regions, the first four of which are labeled according to the number of pulses in transit at any one time.

6.4.4 Cross-Range Resolution

We have discussed the cross-range (along-track) sampling requirements for SAR. Now we turn to finding the cross-range resolution achievable by the radar. We will later make the connection to our understanding of basic signal processing, finding that the cross-range resolution is inversely proportional to the spatial bandwidth observed by the radar. For the moment, we can take a more traditional approach for determining the cross-range resolution of the SAR image by treating the synthetic aperture as real aperture of length L_{SA} and employing the usual argument that cross-range resolution is determined by the product of the range to the target and the beamwidth. For a SAR, this virtual beamwidth is $\theta_{SAR} \approx \lambda/2L_{SA}$. The factor of two improvement in the denominator arises because the synthetic aperture is energized piecewise, as compared with a real aperture that transmits and receives on the entire aperture simultaneously. This difference provides the synthetic aperture with additional phase information that makes it twice as effective as a real aperture of the same length. Continuing this argument, we find that the cross-range resolution can be written as

$$\delta x = R\theta_{SAR} \approx R \frac{\lambda}{2L_{SA}} = \frac{\lambda}{2\theta_{int}} \quad (6.26)$$

where θ_{int} is the synthetic aperture integration angle. This formula holds for all SAR modes as long as θ_{int} is interpreted as the relative rotation between the radar and the target. For inverse SAR, both the target and the platform may move at the same time, making the integration angle less straightforward to compute [20,21]. Regardless of mode, angular diversity is the key to SAR resolution in the cross-range dimension.

6.5 | IMAGE RECONSTRUCTION

In this chapter we focus exclusively on the polar format algorithm for image reconstruction for two reasons: First, PFA is the most common technique currently in practice. Second, an understanding of the operation of PFA is extremely useful for understanding the principles of interferometry and coherent change detection that appear in the next chapter.

When describing the polar format algorithm, it's perhaps best to begin with the punchline and work backward: under the tomographic paradigm, a radar using deramp-on-receive measures directly the spectrum of the scene reflectivity—not the reflectivity itself. The complication is that these measurements are made in a polar coordinate system, while the fast Fourier transform (FFT) requires regularly spaced samples on a Cartesian grid. The key functionality of the PFA is the frequency-domain interpolation that takes the measured data from its native polar representation onto a Cartesian form to which the FFT is applied to obtain the SAR image.

6.5.1 The Tomographic Paradigm

The reflected signal observed at any instant of time $t = 2R/c$ comes from the superposition of all reflections from scatterers at a distance R from the radar. The radiated wavefronts are spherical, but we observe that the transmitted energy is not uniformly distributed throughout the entire 4π steradians comprising the sphere enclosing the radar. By design the antenna's directivity pattern concentrates the radiated energy into a small solid angle, effectively illuminating only a limited region of the earth. If the scene diameter D is

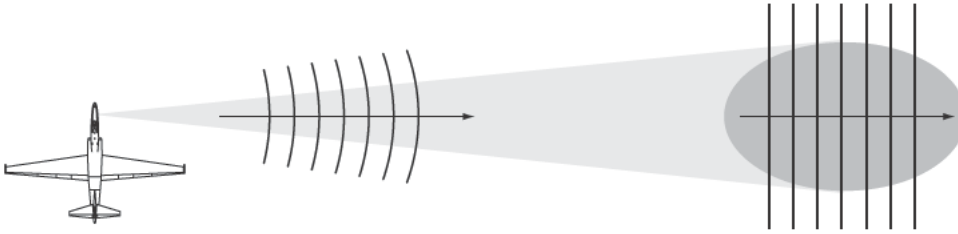


FIGURE 6-10 ■ Assuming the scene size is small relative to the standoff, the incident spherical wavefronts are effectively locally planar. Planar wavefronts permit the use of the projection slice theorem to invert the image from the radar measurements using the Fourier transform.

much less than the standoff distance R , then the wavefront curvature is negligible. The wavefronts are effectively planar over this limited region, as illustrated in Figure 6-10.

It is common in SAR to assume that the scene of interest subtends only a small angle when viewed from the radar. The content of a single received range bin is the projection onto the line of sight of all reflections at that range. Taken together, all of the recorded range bins form the range profile for a single transmitted pulse. Since all reflections from a given range are received simultaneously, and since we assume the wavefronts are planar, the range profile sensed by the radar is the projection of the scene onto a line whose angle is determined by the vector pointing from the scene center to the radar.

The polar format algorithm hinges on a key result known as the projection slice theorem. In words, the projection slice theorem tells us that the following two operations are equivalent for a two-dimensional function $f(x, y)$:

1. Project the values of $f(x, y)$ onto a line ℓ , and then apply the one-dimensional Fourier transform along this line.
2. Given the Fourier transform $F(k_x, k_y) = \mathcal{F}\{f(x, y)\}$, evaluate the Fourier transform along a line (or slice) L that passes through the origin and is parallel to ℓ .

The idea applied to SAR is the same, with two important distinctions: (1) the frequencies measured by the SAR are centered at the radar carrier f_0 and are thus offset from the origin; and (2) the terrain is a three-dimensional function $f(x, y, z)$ so the radar samples its three-dimensional Fourier transform $F(k_x, k_y, k_z)$. The classic paper by Munson and Sanz discusses the problem of reconstructing images using only small portion of the Fourier spectrum offset from the origin [22], and Jakowatz et al. [5] explore the connections between spotlight SAR and X-ray tomography via the projection slice theorem.

6.5.2 The Polar Formatting Algorithm

In spotlight SAR applications each received range profile is Fourier transformed, evaluating a line segment within $F(k_x, k_y, k_z)$. Recall from Section 6.4.1.1 that this transform may be accomplished as a natural by-product of a deramp receiver. Otherwise, the Fourier transform is performed digitally in the subsequent digital processing. As the sensor flies along, the accumulated samples form a surface commonly called the Fourier ribbon. The resulting data set collected by the spotlight SAR is illustrated in Figure 6-11.

The length of the spectral line segment measured by each pulse is equal to the transmitted bandwidth, which is centered at the carrier frequency, shown in Figure 6-11.

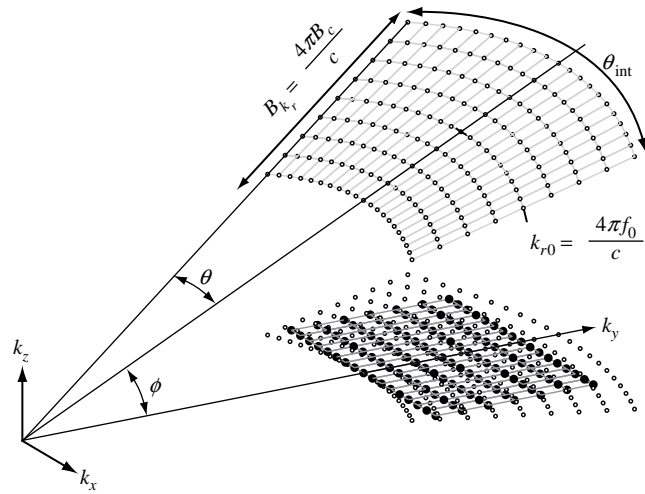


FIGURE 6-11 ■ The size and position of the Fourier ribbon are determined by the radar parameters and collection geometry. The figure shows the center frequency f_0 , transmitted bandwidth B_c , grazing angle ϕ , azimuth angle θ , and integration angle θ_{int} . The polar format algorithm creates the image by projecting the data into the $k_z = 0$ plane, interpolating onto a rectangular grid, and Fourier transforming the result.

These quantities are expressed as range (or radial) wavenumber through the mapping $k_r = 4\pi f/c = 4\pi/\lambda$. Notice that each line passes through the origin in the frequency domain. The angle ϕ is the grazing angle determined by the collection geometry at the time of each transmission. The Fourier ribbon is often idealized as being planar in nature, but variations in platform grazing angle cause it undulate. The set of azimuth angles spanned by the Fourier ribbon is also determined by the collection geometry. The Fourier domain azimuth angle of a given slice is equal to the azimuth of the collected pulse. It follows that the angular spread in the Fourier domain is equal to the integration angle. The angle between pulses can expand and contract as the aircraft speed changes while the radar operates at a fixed PRF.

To reconstruct the image, all that is required is to project the measured 3-D spectrum (the ribbon) into a convenient plane and Fourier transform the result. However, the data must reside on a regularly spaced rectangular grid to exploit the efficiency of the FFT. We may either interpolate the data onto a rectangular grid and then employ the FFT, or we may use the nonuniform discrete Fourier transform directly on the polar-sampled data. Generally, the indirect route of interpolation followed by the FFT is substantially faster than the direct approach.

If we apply the two-dimensional inverse Fourier transform in the plane where the data were collected we will obtain the so-called slant plane image. It has the appearance of the terrain projected along planar wavefronts into a plane oriented at the grazing angle ϕ , as illustrated in Figures 6-4 and 6-5. The primary visual effect is that the image is foreshortened; its range is compressed by a factor of $\cos \phi$. It is possible to interpolate the slant plane image in order to get a geometrically correct ground plane representation. This postprocessing is unnecessary, however. It is more efficient to project the measured $F(k_x, k_y, k_z)$ into the $k_z = 0$ plane as shown in Figure 6-11 and perform the polar-to-rectangular interpolation there. The resulting image will be properly scaled in range. Regardless of slant

plane or ground plane projection, the information content and perspective of the imagery is unchanged. Performing the projection in the Fourier domain is simply a way to avoid the cost of subsequent interpolation in the image domain and the associated potential for introducing phase errors.

Since we have opted to interpolate among the polar values to evaluate $F(k_x, k_y, k_z)$ on a rectangular grid, we must give attention to the types and quality of the interpolators commonly used for the PFA. Better interpolators are typically more computationally expensive and improve the fidelity of the result. As we see in Figure 6-11, the polar-to-rectangular interpolation to be performed appears to require a two-dimensional interpolator. However, the figure is somewhat exaggerated for effect. Most spotlight SAR is rather narrowband, meaning that the polar sampling is actually quite close to the desired rectangular grid.

The two-dimensional polar-to-rectangular interpolation is considered to be separable, meaning that it can be accomplished in two stages using a pair of one-dimensional interpolations [4,5]. The impetus for this is the fact that interpolators are far simpler to implement and analyze in one-dimension than in higher dimensions. The standard procedure is to interpolate first along the radials (varying fast-time frequency) and then along lines of constant fast-time frequency (varying angle). The intermediate result after the first interpolation has the appearance of a keystone, as shown in Figure 6-12(b). After the second interpolation, the data reside on a rectangular grid. Moreover, it is standard practice to extract the largest rectangle that can be inscribed within the polar region of support. This procedure ensures that the point scatterer response (PSR) is spatially invariant within the image. This extracted region of support is usually weighted by a windowing function to control the sidelobe level in the image domain. There are many popular windowing functions [12], but the Taylor window is the most commonly used in SAR.

Figures 6-13 and 6-14 show the properties of the spectral region of support after polar format interpolation. Figure 6-13 shows the region of support and the sampling within it. Recalling that the units of spatial frequency are rad/m, we find the nominal cross-range and range resolution of the image to be $2\pi/B_{k_x}$ and $2\pi/B_{k_y}$, respectively. These will be slightly coarser than the maximum ideal resolution due to the trimming used to obtain a rectangular region of support and any windowing applied in the spectral domain. Assuming that the integration angle is small, at the center frequency the cross-range spatial bandwidth is $B_{k_x} = 4k_0 \sin(\theta_{int}/2) \approx 4\pi f_0 \theta_{int}/c$. The cross-range resolution is then $\delta x = 2\pi/B_{k_x} = \lambda/2\theta_{int}$, agreeing with the earlier result in Section 6.4.4 obtained using a different argument. Meanwhile, the sample spacing in the Fourier domain dictates the area

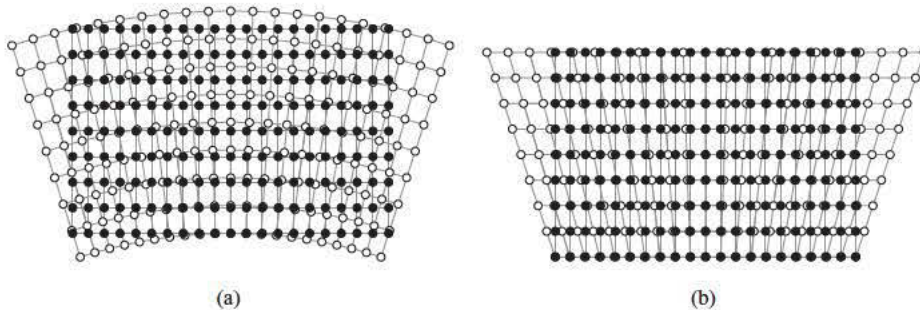


FIGURE 6-12 ■ Interpolation from polar samples (open circles) onto a Cartesian grid (solid circles) is usually done using a two-stage procedure, interpolating first along radials (mapping k_r to k_y) and then along angle (mapping θ to k_x).

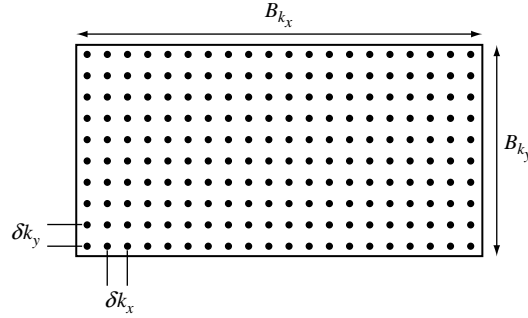


FIGURE 6-13 ■ After the polar format interpolation step, the spatial frequency bandwidth, B_{k_x} and B_{k_y} , determine the image resolution in cross-range and range. Meanwhile, the spectral sample spacing, δk_x and δk_y , determine the image coverage in cross-range and range.

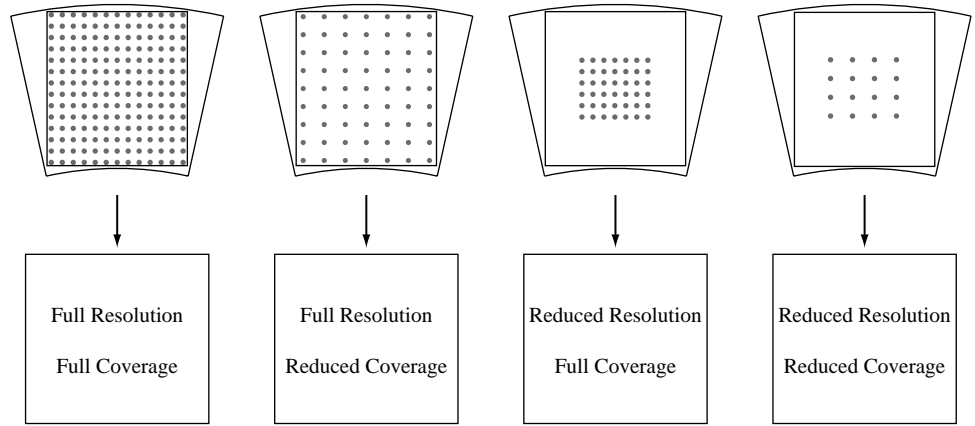


FIGURE 6-14 ■ Diagrams indicating the relationships between spectral support area and sampling and the corresponding image after the 2-D Fourier transform. The annular regions in the top row represent the collected SAR data, while the inscribed rectangles show the maximum region of support that will yield a spatially-variant PSR. (Figure adapted from Jakowatz et al. [5].)

coverage in the final image according to $2\pi/\delta k_x$ and $2\pi/\delta k_y$. Naturally, it doesn't do any good to interpolate onto a grid that is so finely spaced that the image coverage exceeds that of the radar's beam footprint. Such imagery would show the edges of the scene roll off in amplitude according to the beampattern. Figure 6-14 summarizes the effects in the image domain of changing the size of the region of support and the spectral sample spacing.

When implementing the polar format algorithm, the user is free to choose any interpolation scheme whatsoever, but the community has largely settled on the weighted sinc-based interpolator [4,5]. The sinc interpolator has two advantages: its properties are well understood in terms of signal processing metrics, and its performance is easy to dial in so as to trade between speed and accuracy. The truncated sinc interpolator is given by

$$f(x_{out}) \approx \frac{\Delta x_{in}}{\Delta x_{out}} \sum_n f(x_n) w(x_{out} - x_n) \frac{\sin(\pi(x_{out} - x_n)/\Delta x_{out})}{\pi(x_{out} - x_n)/\Delta x_{out}} \quad (6.27)$$

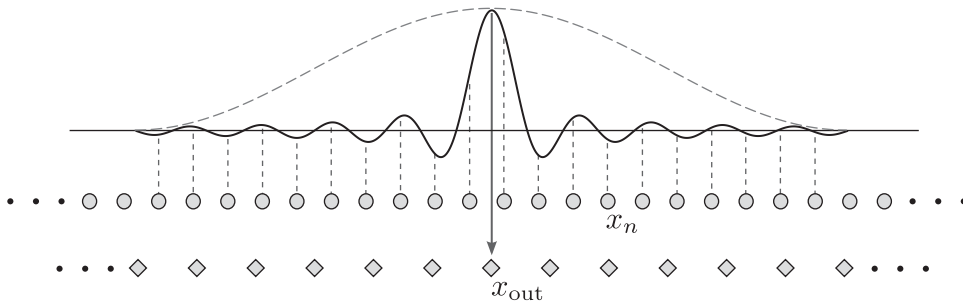


FIGURE 6-15 ■ The truncated sinc interpolator operates by centering the sinc kernel (black curve) on the desired output sample location (diamonds) and then summing all of the input samples $f(x_n)$ that fall under the kernel after first weighting them by the corresponding value of the kernel. It is common practice to apply a weighting function to the sinc kernel (dashed curve).

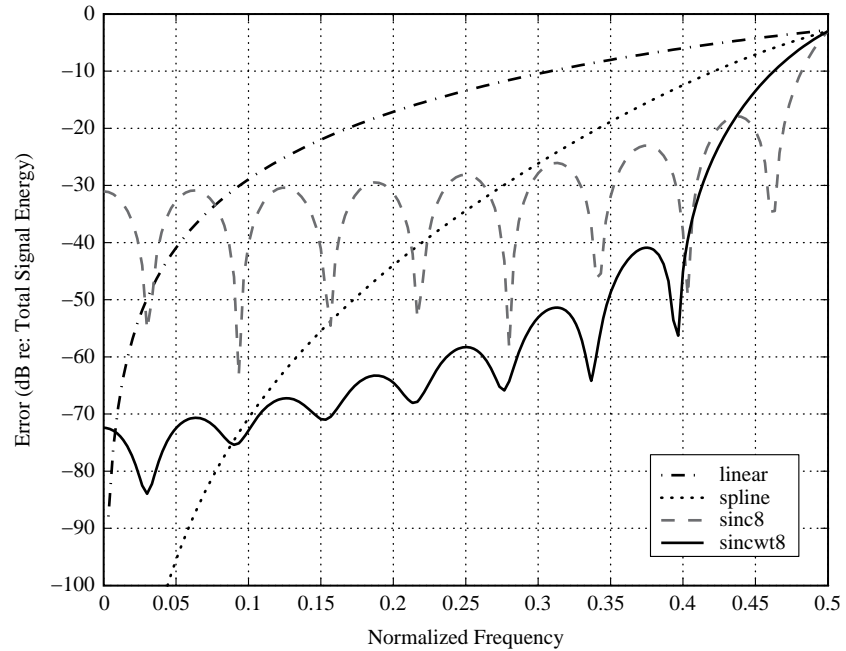
and is shown in Figure 6-15. Each interpolated output $f(x_{out})$ is computed by (1) centering a truncated sinc function at that location, (2) weighting the input samples $f(x_n)$ by the values of the overlapping sinc function, and (3) summing the result. Often, a weighting, or windowing, function $w(x)$ is applied to the sinc kernel to control the level of aliasing in the interpolated output. The quality of the interpolation improves as the length of the truncated sinc grows. It is general practice to specify the interpolation by indicating the number of zero crossings on one side of the truncated sinc. Figure 6-15 shows a 10-point sinc interpolator kernel, and in the diagram the interpolation kernel covers six input samples. It cannot always be assumed that the input sample locations will be uniformly spaced. For example, the angle between pulses (the second interpolation stage in PFA) is unlikely to be uniform because radars typically operate at a fixed PRF during the collection, while the aircraft may change speed over this time due to wind conditions.

Figure 6-16 shows the performance of several common types of interpolators. The horizontal axis shows the normalized frequency of the sinusoid being interpolated, where 0.5 corresponds to one-half of the sampling rate. Thus, at $\omega = 0.5$ the sinusoid passes through π radians between samples, exactly satisfying the Nyquist criterion. The vertical axis represents the summed magnitude-squared interpolation error divided by the total input signal energy. The plot compares linear and piecewise cubic spline interpolation to unweighted and weighted truncated sinc interpolators, both with eight single-sided zero crossings. Recall that the function being interpolated is a superposition of complex sinusoids of various frequencies and amplitudes, where the frequency is proportional to the distance from scene center. A few examples were shown in Figure 6-2. For a fixed sampling rate, higher frequency sinusoids are more difficult to interpolate accurately. Thus, one result of using a better interpolator is to improve the image quality toward the scene edges.

6.5.3 Other Reconstruction Algorithms

While we have emphasized PFA as being the primary reconstruction technique for SAR, its continued dominance is far from ensured. The algorithm was born in an era when computing resources were scarce, and the limitations of PFA were seen as a fair trade for its efficiency. Remember that the FFT was hailed as being the most influential algorithm of the 20th Century. Moving forward, more demanding SAR applications will likely outstrip

FIGURE 6-16 ■
Comparison of
common
interpolators. For
spotlight imagery,
the normalized
frequency
corresponds to
distance from scene
center.



the utility of PFA. For example, wide-beam imaging violates the tomographic assumption. This tends to occur for radars with short standoff ranges and for radars operating at low frequencies, such as those used for foliage penetration. Furthermore, high performance computing hardware and software architectures are increasingly accessible, making brute force backprojection a viable alternative for image reconstruction. For these reasons we must be careful to approach new problems afresh, asking which processing chain is best suited to the given constraints. Nevertheless, the tomographic paradigm is useful for understanding the fundamentals of SAR in the same way that truncated Taylor series expansions are useful for understanding the behavior of nonlinear functions.

6.6 | IMAGE METRICS

There is no single scalar-valued quantity that can reliably indicate if a SAR image is ‘good’, and the particular metric used depends largely on the application. For this reason there are a multitude of ways to judge the quality of a SAR image. When discussing the quality of detail of visible features in the scene, one might apply an image rating scale. Interpretation relies on the magnitude of the image, so the quality of the phase information is of no concern. However, interferometry and coherent change detection work well only if the phase information is preserved with high fidelity from one image to the next. Thus, useful metrics for these applications are formed by computing the cross-correlation as a function of position within the scene. Other metrics are required when designing a SAR or performing trade studies and the goal is to quantify the best theoretical image quality achievable by the radar. Resolution is an obvious choice, but it captures only part of the story. A better metric is the image contrast, which incorporates contributions from the radar hardware, its settings, the environment, and the processing chain [23].

The contrast ratio quantifies the shadow depth achievable relative to a terrain of a given mean radar reflectivity. Equivalently, it determines how bright a no-return area will appear in the image. The ultimate contrast achieved within a SAR image can be limited by the hardware (e.g., antenna dimensions and array configuration, electronic noise floor, and quantization noise), the processing chain, the collection geometry, or the terrain itself. Many of these can be controlled to some extent by the system designer. By studying the contrast ratio, it is possible to balance system parameters to achieve a design that meets the specified requirement.

Carrara et al. [4] describe the distributed target contrast ratio (DTCR) to quantify the ability to visually distinguish between two adjacent types of terrain, one having high backscatter and the other having low. The DTCR is defined as

$$\begin{aligned}
 \text{DTCR} &= \frac{\sigma_{0,\text{high}} + \sigma_{n,\text{total}}}{\sigma_{0,\text{low}} + \sigma_{n,\text{total}}} \\
 &= \frac{\sigma_{0,\text{high}} + \sigma_{n,\text{add}} + \sigma_{n,\text{mult}}}{\sigma_{0,\text{low}} + \sigma_{n,\text{add}} + \sigma_{n,\text{mult}}} \\
 &= \frac{\sigma_{0,\text{high}} + \sigma_n + \text{MNR } \bar{\sigma}_0}{\sigma_{0,\text{low}} + \sigma_n + \text{MNR } \bar{\sigma}_0} \quad (6.28)
 \end{aligned}$$

where the quantities used in this equation are given in Table 6-3. The total noise $\sigma_{n,\text{total}}$ is decomposed into $\sigma_{n,\text{add}}$ and $\sigma_{n,\text{mult}}$, which are the additive and multiplicative noise, respectively. The third line of equation (6.28) comes from [4] where common usage assumes that the most significant multiplicative noise sources can be combined into a single multiplicative noise ratio (MNR) value that is proportional to the average scene backscatter. Special applications might consider additional sources of multiplicative noise, such as volume scattering due to rain. In this case, the multiplicative noise cannot be simplified and written as $\sigma_{n,\text{mult}} = \text{MNR } \bar{\sigma}_0$. The following sections describe the constituent sources of additive and multiplicative noise relevant to SAR.

A typical radar application for the DTCR might compare the (low) backscatter from a smooth asphalt road to the (high) backscatter from surrounding vegetated terrain. We will pursue the limiting case for which the low backscatter regions of interest are shadows or no-return areas, so $\sigma_{0,\text{low}}$ is zero by definition. For notational simplicity the high backscatter region will be denoted simply as σ_0 . This leads to the quantity sought; the terrain to shadow contrast ratio, or more simply, the contrast ratio (CR) of the image,

$$\text{CR} = \frac{\sigma_0 + \sigma_n + \text{MNR } \bar{\sigma}_0}{\sigma_n + \text{MNR } \bar{\sigma}_0} \quad (6.29)$$

TABLE 6-3 ■ Quantities Contributing to DTCR

$\sigma_{0,\text{high}}$	Backscatter coefficient for brighter image region
$\sigma_{0,\text{low}}$	Backscatter coefficient for darker image region
$\sigma_{n,\text{total}}$	Equivalent backscatter coefficient of total noise
$\sigma_{n,\text{add}}$	Equivalent backscatter coefficient of additive noise
$\sigma_{n,\text{mult}}$	Equivalent backscatter coefficient of multiplicative noise
$\bar{\sigma}_0$	Average backscatter coefficient for the terrain
MNR	Multiplicative noise ratio

The value of σ_0 is understood to represent the lowest backscatter in the scene, which is to be distinguished from a shadow. Note that equation (6.29) compares backscatter coefficients, so it is necessary to express the various noise contributions in these terms.

Two values of the contrast ratio are of particular interest. First, note that CR approaches 1 (0 dB) as the backscatter coefficient goes to zero. The CR can approach this case, depending on the terrain type, imaging geometry, and the radar frequency. A second interesting case occurs when the total noise contribution is equal to the terrain scattering, resulting in $CR = 2$ (3 dB). This is a reasonable starting point for determining minimum acceptable contrast requirements, although contrast ratios of 15 dB or better are required for obtaining high-quality imagery.

6.6.1 Additive and Multiplicative Noise

Additive noise is that form of noise that would be present in the recorded signal even with the transmitter deactivated. The most significant contribution is thermal noise:

$$\sigma_{n,add} = \sigma_{thermal} \quad (6.30)$$

We require the additive noise to be expressed as an effective backscatter coefficient that may be derived from a signal-to-noise metric known as the clutter-to-noise ratio (CNR): $CNR = \sigma_0/\sigma_{n,add}$. The clutter (desired signal) portion of CNR is the specific backscatter coefficient of the terrain of interest. The noise component is the thermal noise expressed as a specific backscatter coefficient, $\sigma_{n,add}$ (or simply σ_n), which is also often called the noise-equivalent sigma zero (NESZ). It is the effective value of σ_0 that would produce a received signal equal in power to the system noise. The additive noise term needed for (6.29) is obtained from $\sigma_{n,add} = \sigma_0/CNR$. The expression for CNR is found in several sources, such as [3,4] and equation (21.79) in *Basic Principles* [14].

Multiplicative noise is so named because it rises and falls in proportion to the backscattered signal. The multiplicative noise ratio is independent of the transmitted signal power because doubling the transmitted power doubles both the backscattered signal and the multiplicative noise power. Sidelobes of the point scatterer response are a good example of a multiplicative noise source: increasing the transmitted power increases both the main lobe and the sidelobes by the same fraction.

According to [4] the MNR is approximated by summing the primary sources of multiplicative noise, which are the range and along-track ambiguity-to-signal ratios (RASR and AASR), the integrated sidelobe ratio (ISLR) of the image impulse response, and the quantization noise ratio (QNR). These are combined as

$$\sigma_{n,mult} = \bar{\sigma}_0(ASR + ISLR + QNR) \quad (6.31)$$

The sources of multiplicative noise considered in this chapter are described below. Others can be found in Table 8.10 of [4].

6.6.1.1 Ambiguity-to-Signal Ratio

The ambiguity to signal ratio (ASR) represents the superposition of along-track ambiguities (also called azimuth or Doppler ambiguities) and range ambiguities within the imaged scene. The primary assumption made is that all of the ambiguities are uncorrelated, which allows their total contribution to be represented by a simple summation. The ASR is a scale factor and has no reference quantity associated with its decibel representation.

Along-track ambiguities arise because the synthetic aperture is not sampled finely enough to avoid aliasing all spatial frequencies. The pulses would have to be spaced at or less than $\lambda/4$ to avoid spatial aliasing altogether. Aliased returns manifest as multiplicative noise in the image, and the typical effect on the image is simply reduced contrast. However, highly-reflective discrete objects can cause visible ghosting in the imagery (e.g., Figure 7-28 in [3]).

Range ambiguities are echoes from earlier or later transmissions superimposed on the return from the current pulse. Whereas the along-track ambiguities are attenuated primarily by the transmit/receive beampattern, the range ambiguities are attenuated by spherical spreading loss, the beampattern of the radar, and the fact that the grazing angle (and thus terrain reflectivity) becomes very small at ranges far beyond the imaged scene. Range ambiguities are also terminated by the radar horizon caused by the earth's curvature.

The total ASR is given by the integral in equation (6.32), where $G = G_{TX}G_{RX}$ is the composite two-way gain for a particular spatial frequency k_x (proportional to the angle from boresight for a broadside-looking sensor), σ_0 is the backscatter coefficient of the terrain, B_p is the processed spatial bandwidth, k_{xs} is the spatial sampling frequency in rad/m ($k_{xs} = 2\pi/\delta x$), δx is the distance between along-track samples, and f_p is the PRF in Hz. The ASR is expressed as a function of time τ (or equivalently, range $R = c\tau/2$):

$$\text{ASR}(\tau) = \frac{\sum_{\substack{m,n=-\infty \\ m,n \neq 0}}^{\infty} \int_{-B_p/2}^{-B_p/2} G(k_x + mk_{xs}, \tau + n/f_p) \cdot \sigma_0(k_x + mk_{xs}, \tau + n/f_p) dk_x}{\int_{-B_p/2}^{-B_p/2} G(k_x, \tau) \cdot \sigma_0(k_x, \tau) dk_x} \quad (6.32)$$

The literature frequently describes the along-track ambiguities in terms of Doppler frequency, rather than spatial frequency, as is done here. The formulations are interchangeable, and the two are related by $f_D = vk_x/2\pi$, where v is the forward speed of the sensor. For a fixed temporal frequency, the spatial frequency is proportional to the angle of arrival θ of the reflected signal: $k_x = 2k \sin \theta$ [17,24].

The processed bandwidth is the extent of the spectral region of support of the wavenumber k_x (recall Figures 6-13 and 6-14). Thus, the processed bandwidth is equivalent to an effective beamwidth. It is bounded above by the composite transmit/receive beamwidth, and it may be intentionally reduced to achieve better ASR performance at the expense of a smaller scene size. Interestingly, the processed spatial bandwidth has different meanings for spotlight and stripmap imagery. The common thread is that reducing processed bandwidth equates to narrowing the effective beamwidth. For a fixed wavelength λ , narrower beams reduce the length of the synthetic aperture (thus coarsening resolution) in stripmap mode while they reduce the size of the illuminated scene in spotlight mode.

6.6.1.2 Integrated Sidelobe Ratio

The integrated sidelobe ratio is the ratio of the energy in the sidelobes of the image impulse response to the energy in its main lobe. Specifically the quantity of interest is the two-dimensional ISLR. Like the ASR, it is a scale factor and has no reference quantity associated with its decibel representation. Recalling that multiplication in one Fourier domain is convolution in the other, we see that the spectral trimming and windowing applied as part of the PFA reconstruction image determines the impulse response. The ISLR values for several common windows are given in Table 6-4.

TABLE 6-4 ■ Properties of Common 2-D Spectral Weighting Functions

2D Window	ISLR (dB)	Main Lobe Broadening
Rectangular	−6.5	1
Taylor (5/35)	−24.6	1.32
Hamming	−31.3	1.46
Hanning	−29.9	1.56

Note: The main lobe broadening factor is relative to the main lobe width of the uniform window (sinc) response.

Under ideal conditions, the ISLR of the image impulse response is a function of the spectral support and weighting associated with the image reconstruction. However, phase errors can corrupt the impulse response causing the ISLR to deviate from the values given in Table 6-4. The impact of phase errors on the impulse response is described in Chapter 5 of [4], [25], and in Section 6.7 below.

6.6.1.3 Quantization Noise

A number of models can be used to describe quantization noise introduced by the ADC [26,27]. The QNR of an ideal converter is $-(6.02N + 1.76)$ dB, where N is the number of bits. Another simple rule-of-thumb found in [4] is that QNR is given by -5 dB per bit used in the ADC. This QNR is somewhat conservative relative to other measures. Like ASR and ISLR, the QNR is a scale factor and has no reference quantity associated with its decibel representation.

6.6.2 Rating Image Quality

SAR image quality is quantified in a number of ways, some of which are subjective. Among these is the National Image Interpretability Rating Scale for radar imagery (radar NIIRS), or RNIIRS for short. The ratings are based on the usefulness of the imagery for carrying out specific types of tasks. The RNIIRS system assigns to an image a number ranging from 0 to 9, with 9 being the highest quality. The RNIIRS value gives an indication of the information content of the imagery. While the RNIIRS system takes into account the resolution (also called the ground sampled distance), it is important to observe that resolution is not necessarily tightly coupled to the RNIIRS rating. For example, a SAR capable of achieving 0.3 m resolution might seem to have a rating of 7. However, the image could be degraded by an unfavorable imaging geometry, phase errors, or noise such that the actual rating is much lower. The RNIIRS system is described in [28] and is summarized in Table 6-5.

6.7 | PHASE ERROR EFFECTS

Our development of spotlight SAR has been centered on the notion that each scatterer in an image corresponds to a single frequency (or linear phase progression) in the Fourier domain. Deviations from this idealization can come from a number of sources and are known as phase errors because they are usually assumed to be equivalent to distances that are smaller than a range bin. The phase error effects are then treated as being one-dimensional,

TABLE 6-5 ■ Image Resolution and Example Tasks Used to Define RNIIRS.

Rating	Resolution	Tasks
0	N/A	Image unusable due to hardware failure or interference.
1	>9 m	Detect a large cleared swath in a densely wooded area. Detect lines of transportation (either road or rail), but do not distinguish between them.
2	4.5–9 m	Detect the presence of large bombers or transports. Detect a military installation by building pattern and site configuration. Identify athletic stadiums.
3	2.5–4.5 m	Detect medium-sized aircraft. Detect vehicle revetments at a ground forces facility. Identify a medium-sized railroad classification yard.
4	1.2–4.5 m	Distinguish between large rotary-wing and medium fixed-wing aircraft. Detect individual vehicles in a row at a known motor pool. Detect all rail/road bridges.
5	0.75–1.2 m	Count medium helicopters. Distinguish between river crossing equipment and medium/ heavy armored vehicles by size and shape. Count railcars by detecting the breaks between them.
6	0.4–0.75 m	Distinguish between variable and fixed-wing fighter aircraft. Distinguish between small support vehicles and tanks.
7	0.2–0.4 m	Identify small fighter aircraft by type. Detect street lamps in an urban residential area or military complex.
8	0.1–0.2 m	Distinguish the fuselage difference between similar aircraft. Identify the dome/vent pattern on rail tank cars.
9	<0.1 m	Detect major modifications to large aircraft (e.g., fairings, pods, winglets). Identify trucks as cab-over-engine or engine-in-front.

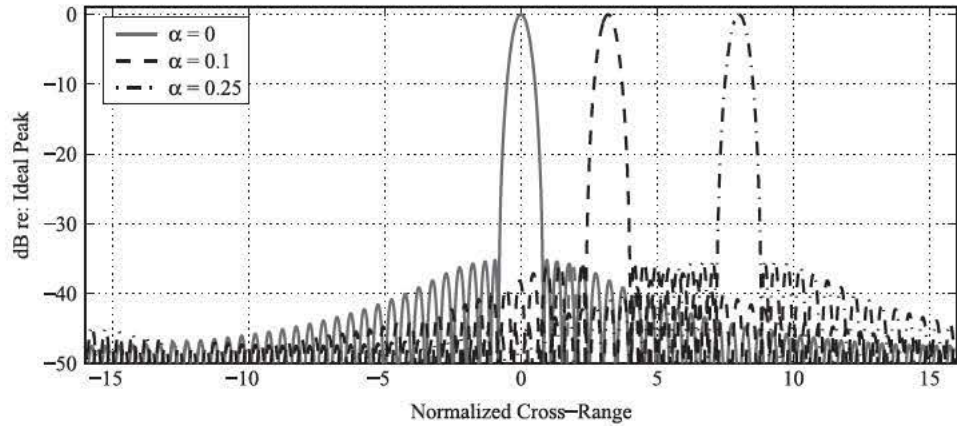
(Table adapted from [28].)

as they are constrained to the cross-range axis, varying only with slow time. If the radar system is well designed the primary sources of error are platform motion measurement error, target motion, and approximations in the signal processing chain.

To quantify the impact of phase errors on the SAR imagery, we must first understand the mechanism by which the image is corrupted. We model the phase error as a function $g(\theta)$ of angle (or slow time or pulse number) that multiplies each received pulse. Under the tomographic paradigm the SAR data represent measurements of the Fourier transform of the scene, so the corrupted image domain point scatterer response (PSR) is the convolution of the ideal PSR with the Fourier transform of the phase error function [4,5]. We will see in the next section that this property is successfully exploited by autofocus algorithms to restore image quality. Since $g(\theta)$ does not depend on range or temporal frequency ω it is constant for each pulse and can be factored out of the Fourier transform of a single pulse. Therefore, it makes no difference if the SAR employs a deramp receiver or samples the scattered signal directly.

Regardless of their origin, phase errors usually fall into several common categories and are characterized by the frequency of the error relative to the length of the synthetic aperture. For example, $\sin(\theta)$ can appear to be linear if it is observed for only a short interval near the zero crossings $\theta = 2\pi m$ or quadratic if observed near the extrema $\theta = \pi/2 + 2\pi m$. If observed for a longer time, the oscillatory nature of $\sin(\theta)$ becomes evident. If $g(\theta)$ is random in nature, it is said to be wideband. These four types of phase error—linear, quadratic, sinusoidal, and wideband—are the most commonly seen in SAR imagery, and we examine each in detail. A more thorough treatment is given in Chapter 5 of [4]. It is common to treat the phase error as a property of the synthetic aperture and independent of the scene itself. It is also possible for an individual scatterer to induce phase errors relative to its own ideal phase history, typically because it is moving. In this event,

FIGURE 6-17 ■
Point scatterer
response as affected
by linear phase
errors with varying
slope α .



the surrounding terrain will be well focused, with only a very local region exhibiting the effects of the phase error.

6.7.1 Linear Phase Errors

As stated earlier, only the Fourier transform of $g(\theta)$, $G(x)$, is needed in order to understand the impact of the phase error. A linear phase error is just a complex sinusoid, represented as $g(\theta) = \exp\{-j2\pi\alpha\theta\}$. We saw earlier that its Fourier transform is a delta function whose displacement from the origin is proportional to the frequency α . Since the corrupted PSR is the cross-range convolution of the ideal PSR and $G(x)$, we find that the effect of a linear phase error is a cross-range shift of the image and is generally not accompanied by any defocus of the PSR. However, very large linear phase errors (or Doppler shifts) can move the cross-range SAR signal out of the processing bandwidth, resulting in a loss of resolution. The effect of linear phase error is shown in Figure 6-17. In this and the following three examples of phase errors, the PSR has had a Taylor window applied. The cross-range axis is normalized such that the null-to-null width of the unweighted PSR is equal to 1. Thus, the null-to-null width of the ideal Taylor-weighted PSR is slightly larger than 1. Furthermore, the integration angle θ_{int} is normalized to be 1.

Linear phase errors can occur at the sensor level if the line of sight component of the sensor velocity is not measured accurately. Furthermore, a moving object introduces a linear phase term on its own response if its velocity has any component along the line-of-sight to the radar. In this case, the resulting linear phase will cause the object to be displaced in the cross-range dimension. This effect famously results in cars appearing off of the roads they are traveling on.

6.7.2 Quadratic Phase Errors

The quadratic phase error is given by $g(\theta) = \exp\{-j2\pi\beta\theta^2\}$. There is no closed-form expression for the Fourier transform of a complex exponential with quadratic phase. The solution is instead expressed in terms of Fresnel integrals that must be computed numerically. Alternatively, we can apply the principle of stationary phase [29–31] to obtain a closed-form expression that approximates the desired Fourier transform. Taking this approach we find that, in general, the Fourier transform of a quadratic-phase complex

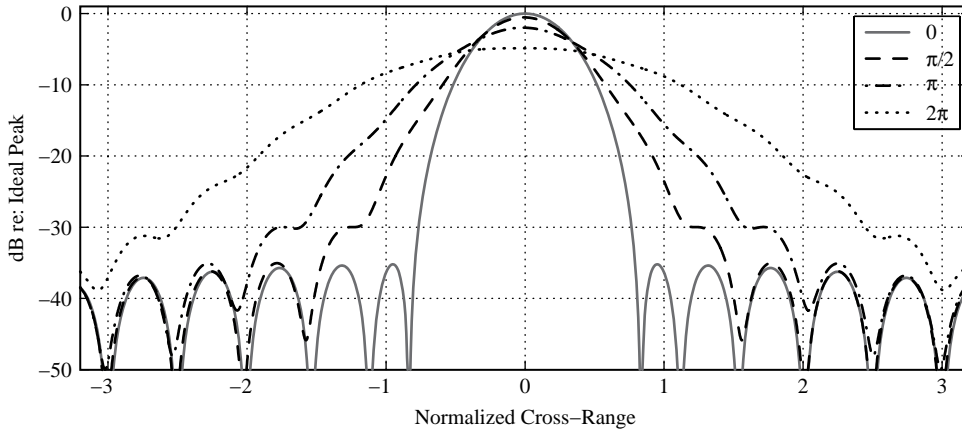


FIGURE 6-18 ■ Point scatterer response as affected by various amounts quadratic phase error corresponding to $\beta = 0, 1, 2$, and 4.

exponential is another quadratic-phase complex exponential. In the image domain, the width of $G(x)$ is found to be $4\pi\beta\theta_{int}$, where θ_{int} is the integration angle of the collection. Furthermore, we see that the width of the resulting PSR is determined by the accumulated phase of $g(\theta)$. Thus, larger values of β and longer synthetic apertures serve to widen $G(x)$. Approximating $G(x)$ as a rect function, we find that the quadratic phase error smears the cross-range PSR causing the image to be defocused. This effect is shown in Figure 6-18 where the PSR is corrupted by quadratic phase errors equal to $2\pi\beta\theta_{int}^2/4$ for several values of β . The QPE is characterized by its maximum value, found at the edges of the synthetic aperture, where $\theta = \pm\theta_{int}/2$ and $\theta_{int} = 1$.

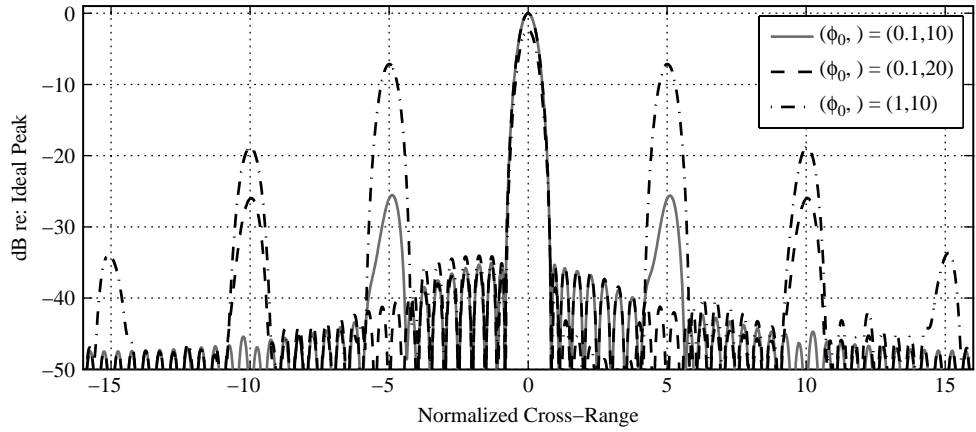
The quadratic phase error is important for analyzing SAR performance because it can be used to model a number of significant effects. For example, it can be used to predict the consequences of using the tomographic paradigm when the radiated wavefronts are in fact spherical. The QPE can also be used to model the appearance of moving objects in spotlight SAR imagery, as the cross-range component of target velocity and line-of-sight acceleration both create additional quadratic phase on top of the ideal response. If the QPE is small, the moving object will be mildly defocused. If it is large, the object can be smeared across many cross-range bins. Especially severe QPE can spread the energy of the moving object so much that it falls below the reflectivity of the surrounding terrain. In such cases, the moving object is invisible in the image, but may be detected by comparison with another image of the same scene using coherent change detection techniques.

6.7.3 Sinusoidal Phase Errors

The sinusoidal phase error $g(\theta) = \exp\{j\phi_0 \sin(2\pi\gamma\theta)\}$ is the third type of phase error we discuss, where ϕ_0 is the zero-to-peak amplitude of the phase error, and γ is its frequency. Our analysis is based largely on the work by Fornaro [32]. The function $g(x)$ can be manipulated into a more useful form by employing the generating function for the Bessel function of the first kind:

$$\exp\left\{\frac{\xi}{2}\left(t - \frac{1}{t}\right)\right\} = \sum_{n=-\infty}^{\infty} t^n J_n(\xi) \quad (6.33)$$

FIGURE 6-19 ■ Sinusoidal phase error effects for several values of amplitude ϕ_0 and frequency γ .



and making the substitutions $t = \exp\{j2\pi\gamma\theta\}$ and $\xi = \phi_0$. The result is:

$$\exp\{j\phi_0 \sin(2\pi\gamma\theta)\} = \sum_{n=-\infty}^{\infty} (e^{j2\pi\gamma\theta})^n J_n(\phi_0) \quad (6.34)$$

Taking the Fourier transform of this with respect to θ gives $G(x)$:

$$G(x) = \sum_{n=-\infty}^{\infty} J_n(\phi_0) \delta(\theta - 2\pi n\gamma) \quad (6.35)$$

Thus, a sinusoidal phase error will result in the creation of replicas of the ideal PSR. These replicas are located at $x = \pm 2\pi n\gamma$ and are weighted according to the value of the coefficient $J_n(\phi_0)$. Although there are an infinite number of replicas, the coefficients die off quickly. The rate of decay slows with increasing sinusoidal amplitude ϕ_0 , and the replica spacing is determined by the frequency γ of the phase error function. Figure 6-19 shows $G(x)$ for representative values of ϕ_0 and γ . A similar result is derived in [4], but only the terms for $n = -1, 0$, and $+1$ in (6.34) are retained on the grounds that the zero-to-peak error amplitude ϕ_0 is small.

Sinusoidal phase errors can arise when the SAR antenna experiences periodic vibration or, perhaps more commonly, when objects in the scene rotate or vibrate. An example of the latter is air-handling equipment on the roof of a building. Sinusoidal phase errors are often obvious in SAR imagery because they result in paired echoes or cross-range ghosts around the vibrating object. If the error occurs at the sensor, the paired echoes will appear over the entire scene resulting in duplicate bridges, roads, buildings, and so on.

6.7.4 Wideband Phase Errors

Wideband phase errors model effects such as antenna vibration due to turbulent air flow past the aircraft, oscillator phase noise, as well as other mechanical and electronic stability issues. The relevant error function is $g(\theta) = \exp\{j\kappa(\theta)\}$, where $\kappa(\theta)$ is a Gaussian random process with variance σ_κ^2 . The effects of a random phase error are shown in Figure 6-20. We see in the figure—and prove in this section—that random phase errors increase the average sidelobe level while leaving the main lobe unaffected.

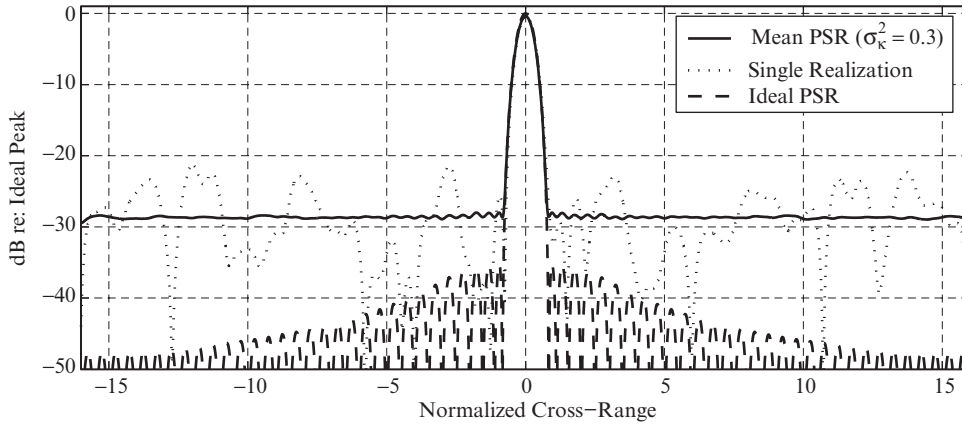


FIGURE 6-20 ■ The expected PSR contaminated by a phase error whose variance is 0.3 radians (solid line), a single realization (dotted line), and the ideal PSR (dashed line).

The average effect on the PSR is found by taking the expected value of the magnitude-squared of the corrupted PSR, $\mathcal{E}\{|F_c(x)|^2\}$. Taking the Fourier transform gives:

$$\begin{aligned}
 \mathcal{E}\{|F_c(x)|^2\} &= \mathcal{E}\{F_c(x) \cdot \bar{F}_c(x)\} \\
 &= \mathcal{E}\{\mathcal{F}\{f_c(\theta) * \bar{f}_c(-\theta)\}\} \\
 &= \mathcal{E}\{\mathcal{F}\{[f(\theta)g(\theta)] * [\bar{f}(\theta)\bar{g}(\theta)]\}\} \\
 &= \mathcal{E}\{\mathcal{F}\{[\text{rect}(\theta/\theta_{int})g(\theta)] * [\text{rect}(\theta/\theta_{int})\bar{g}(\theta)]\}\} \\
 &= \mathcal{E}\{\mathcal{F}\theta_{int} \cdot \{\text{triangle}(\theta/\theta_{int}) \cdot [g(\theta) * \bar{g}(\theta)]\}\} \\
 &= \mathcal{F}\theta_{int} \cdot \{\text{triangle}(\theta/\theta_{int}) \cdot \mathcal{E}\{[g(\theta) * \bar{g}(\theta)]\}\} \\
 &= \mathcal{F}\theta_{int} \cdot \{\text{triangle}(\theta/\theta_{int}) \cdot R_{gg}(\theta)\} \\
 &= \delta x^{-2} \cdot \text{sinc}^2(x/\delta x)
 \end{aligned} \tag{6.36}$$

where R_{gg} is the autocorrelation of g , S_{gg} is its power spectral density [33], and the scaling between x and θ_{int} is found in Equation 7.26 and the discussion in Section 7.5.2. We thus find that the expected value of the corrupted magnitude-squared PSR is described by the ideal PSR convolved with the power spectral density of the autocorrelation of the phase error function $g(\theta)$.

To find the autocorrelation R_{gg} we employ the fact that $g(\theta)$ can be expressed as the sum of its mean and the centered version of itself, $g(\theta) = \mu_g + (g(\theta) - \mu_g) = \mu_g + v(\theta)$.

$$\begin{aligned}
 R_{gg} &= \mathcal{E}\{g(\theta)\bar{g}(\theta + \zeta)\} \\
 &= \mathcal{E}\{(v(\theta) + \mu_g)(\bar{v}(\theta + \zeta) + \bar{\mu}_g)\} \\
 &= \mathcal{E}\{v(\theta)\bar{v}(\theta + \zeta) + v(\theta)\bar{\mu}_g + \bar{v}(\theta + \zeta)\mu_g + \mu_g\bar{\mu}_g\} \\
 &= \sigma_g^2\delta(\theta) + |\mu_g|^2
 \end{aligned} \tag{6.37}$$

The Fourier transform of this is simply:

$$S_{gg}(x) = \sigma_g^2 + |\mu_g|^2 \delta(x) \tag{6.38}$$

Inserting (6.38) into the last line of (6.36) we find that the mean effect of a random phase error is to scale the squared ideal PSR $|F(x)|^2$ by μ_g while adding background noise equal to the variance of $g(x)$ times the total energy of the ideal PSR $F(x)$,

$$\mathcal{E}\{|F_c(x)|^2\} = \mu_g^2 |F(x)|^2 + \sigma_g^2/\delta x \tag{6.39}$$

To obtain closed-form results, we use expressions for the mean and variance of $g(\theta)$, which are derived in a 2003 paper by Richards [34] to be: $\mu_g = \exp\{-\sigma_\kappa^2/2\}$ and $\sigma_g^2 = 1 - \mu_g^2$. We can use these to find the change in ISLR relative to the ideal ISLR, which is defined as the increased integrated sidelobe energy (SLE) divided by the main lobe energy (MLE).

$$\begin{aligned} \text{ISLR} &\approx \frac{\mu_g^2 \text{SLE}_0 + \sigma_g^2 / \delta x}{\mu_g^2 \text{MLE}_0} \\ &= \text{ISLR}_0 + \frac{\sigma_g^2 / \delta x}{\mu_g^2 \text{MLE}_0} \end{aligned} \quad (6.40)$$

We can simplify this expression further by approximating the main lobe energy as a rectangle whose height is equal to the height of the magnitude-squared main lobe and whose width is the -3 dB resolution. The former is equal to δx^{-2} while the latter is approximated by δx . Thus, we find $\text{MLE}_0 \approx \delta x^{-1}$ and

$$\text{ISLR} = \text{ISLR}_0 + \frac{\sigma_g^2}{\mu_g^2} \quad (6.41)$$

where the second term represents the additional ISLR due to the phase noise term. Substituting in the expressions for the mean and variance of $g(t)$, this is found to be $\Delta \text{ISLR} = \exp\{\sigma_\kappa^2\} - 1$. The corrupted ISLR can be used directly in equation (6.29) to predict the loss of contrast due to the random phase error.

Figure 6-21 shows the effect of the image-degrading phase errors on actual SAR data. The imagery was collected by the U.S. Air Force Research Laboratory [15]. The primary feature in the scene is parking lot. Beside it is a road extending roughly vertically in the image.

6.8 | AUTOFOCUS

We see from the previous section that synthetic aperture radars are highly sensitive to phase errors. For example, we find from Figure 6-18 that a two-way quadratic phase error of $\lambda/2$ can effectively destroy the image, and we would like the two-way error to be less than $\lambda/4$ (or $\lambda/8$ one-way) to maintain good focus. If we assume the radar operates at X band ($\lambda = 3$ cm), we find that we must know the platform motion to less than 4 mm over the collection time, which can be as long as several seconds or a minute.

The phase errors must be reduced to tolerable levels to obtain well-focused imagery, and there are two general methods for doing so. The first line of defense, so to speak, is to estimate the motion of the antenna phase center as accurately as possible using a GPS-aided inertial navigation system (INS). However, even high-performance devices may be incapable of measuring all motions and the resulting SAR image is very likely to contain residual error causing it to be at least slightly defocused.

The second approach for mitigating phase error is to employ autofocus algorithms after the motion compensated data has been used to form an image. The term autofocus implies that these techniques operate directly on the SAR image to estimate phase errors. Generally speaking, autofocus routines require as input only the complex imagery and only basic knowledge of the radar system parameters. Spotlight-mode autofocus algorithms exploit

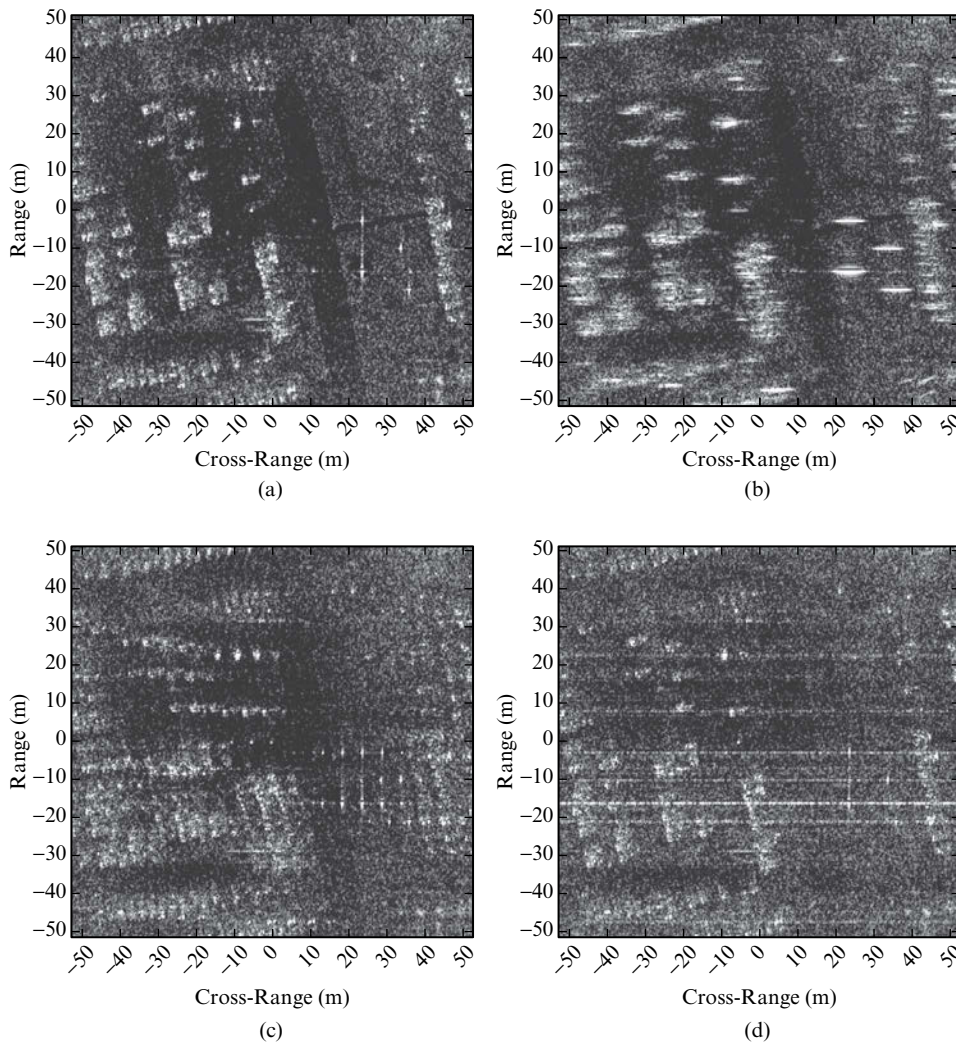
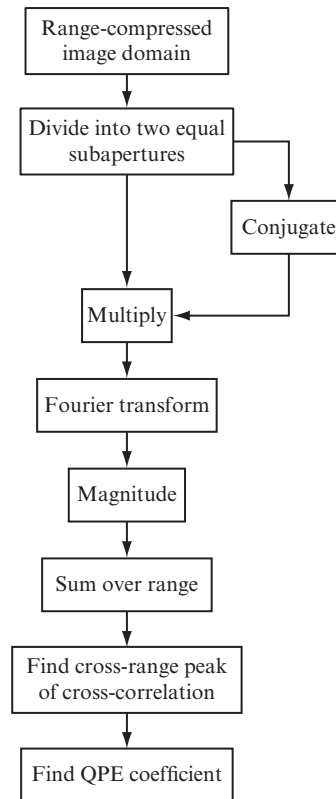


FIGURE 6-21 ■ Example SAR imagery [15] exhibiting phase error effects: (a) No error. (b) Quadratic error. (c) Sinusoidal error. (d) Random error.

the fact that phase errors occurring at the radar result in spatially invariant effects in the corresponding image. This spatial invariance is a strong advantage because it allows a significant amount of averaging to be performed over the scene to obtain a good estimate of the phase error. For cases in which the artifacts are localized, the image can be chipped into small regions and analyzed separately. This is the case for stripmap SAR, for example.

Autofocus techniques are characterized as being either parametric or nonparametric. The former operate by estimating the values associated with a phase error model, such as the quadratic coefficient β or some other polynomial fit, while the latter do not. Parametric techniques are generally good for correcting low-frequency errors, and nonparametric approaches are well-suited to removing random errors. There is some degree of overlap between the two approaches, but it is reasonable to include both in a practical SAR image formation processor since they are complementary. A parametric algorithm is first applied to remove the quadratic phase component and improve focus. This is followed by a nonparametric technique whose purpose is to remove the random phase error to improve image contrast. We now give an example of each type.

FIGURE 6-22 ■
Steps for the phase
difference autofocus
algorithm. (Adapted
from Carrara
et al. [4].)



6.8.1 Phase Difference Autofocus

Phase difference autofocus (PDA) is a noniterative parametric technique for estimating the quadratic phase error present in a SAR image [4]. The procedure is diagrammed in Figure 6-22. If we consider a complex-valued SAR image to be our source of data, we first inverse Fourier transform in the cross-range dimension. This new domain corresponds roughly to the slow-time (or azimuth or pulse) domain of the data and is the input for PDA. The data are then divided into a pair of equal-duration slow-time blocks. The blocks are known as subapertures because their Fourier transforms should yield the same imagery [22], having coarser cross-range resolution and independent speckle realizations.

Each subaperture now contains half of the original quadratic phase. These new phase terms can be approximated by a linear phase plus a residual quadratic phase. The linear component is proportional to the quadratic phase error coefficient β and causes the subaperture-based images to shift in opposite directions. The residual quadratic phase causes each subaperture image to be slightly defocused. PDA operates by cross-correlating the image pair and measuring the offset of the resulting peak to estimate β .

The next step is to multiply the subapertures after having first taken the complex conjugate of one of them. Recall from Table 6-1 that the operation of conjugate multiplication in one Fourier domain is the same as correlation in the other. The data are then Fourier transformed giving the cross-range cross-correlation function evaluated at each range bin. We then take the magnitude of the cross-correlation and then average over all ranges to obtain a single estimate of the subaperture image cross-correlation function. The offset of

the peak of the correlation, Δx , is found and then related to the QPE coefficient via

$$\beta = \frac{\Delta x}{L_{SA}} \quad (6.42)$$

where L_{SA} is the length of the synthetic aperture. The estimated β is used to form the conjugate of the true phase error function, which is in turn applied at each range of the data prior to Fourier transforming to the image domain.

The phase difference algorithm is very similar to another autofocus technique known as map drift (MD) [4,5]. The primary difference between the two is that MD performs the cross-correlation on the real-valued intensity subaperture imagery, whereas PDA cross-correlates the complex-valued imagery. Furthermore, PDA averages over all range bins while MD operates only on those ranges with very bright scatterers, typically a small percentage of the total. This latter difference means that MD is often implemented in an iterative fashion, while PDA can achieve the same result in a single iteration.

6.8.2 Phase Gradient Autofocus

Phase gradient autofocus (PGA) is a widely used technique for the nonparametric estimation of phase errors. It is particularly well suited to the estimation of high-frequency errors, but is capable of dealing with some amount of low-frequency error as well. The PGA algorithm is outlined in Figure 6-23 and begins with the complex image. For each range bin, the cross-range signal is circularly shifted so that the brightest pixel is placed at $x = 0$. The resulting line of bright pixels is windowed in cross-range to extract the corrupted PSR and its sidelobe structure. The remaining cross-range values are set to zero.

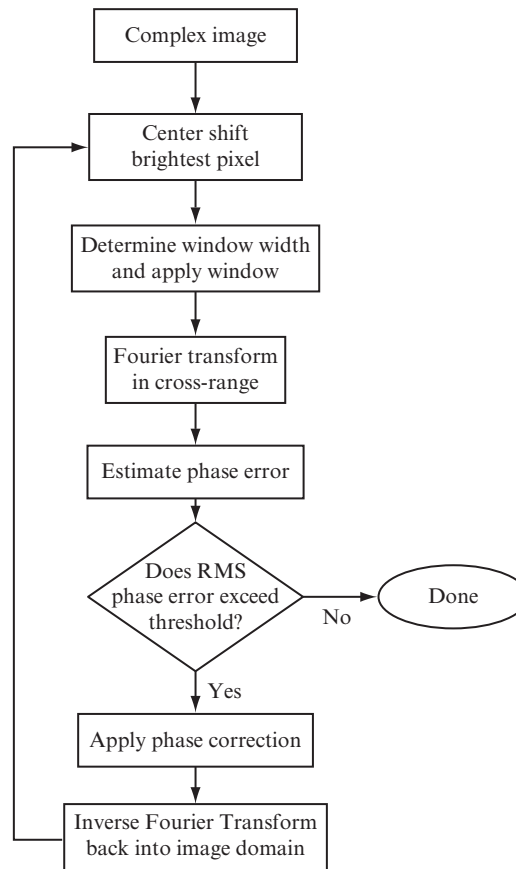
The windowed data are then Fourier transformed in the cross-range dimension. Recall that this operation takes the data from a domain where the ideal PSR is convolved with the phase error spectrum and places it in one where the ideal phase history is multiplied by the phase error function. Since the center of the PSR is close to $x = 0$, its phase history is approximately constant, meaning that any phase terms present are due to noise and the phase error we wish to estimate. For each range, this estimation is accomplished by finite differencing the phase history in the angle (or slow-time) dimension and then re-integrating it. This step eliminates any bias and enables us to average the data over all ranges to obtain the desired phase error estimate.

The conjugate of the estimated error is then multiplied against the data to correct it, and the data are then Fourier transformed back into the image domain. The process is repeated until a stopping criteria is reached. Since PGA targets high-frequency phase errors, a good choice of stopping criteria is the root mean square value of the estimated phase error.

6.9 | SUMMARY AND FURTHER READING

Synthetic aperture radar imaging relies heavily on concepts from Fourier analysis. We introduced spotlight SAR by reviewing a few key properties such as the Fourier transform relationship between the rect and sinc functions as well as the transform of a shifted function $s(t - t_0)$. The property stating that delta functions in one domain Fourier transform into plane waves in the other domain is used to model the terrain as a continuum of

FIGURE 6-23 ■
Steps for the phase
gradient algorithm.
(Adapted from
Jakowatz et al. [5].)



independent point scatterers. These and other basic signal processing concepts were then used to develop the the range and cross-range sampling and resolution constraints for spotlight SAR imaging. We also presented the features of a spotlight-mode SAR image collection, noting that the geometry affects the appearance of the image through the way that ground reflections are projected into the slant plane. An important point to remember is that most spotlight collections employ a linear flight path even though the data are motion compensated to appear as if the flight path were circular.

Spotlight-mode image reconstruction can be carried out using a number of different algorithms. The most popular of these is the polar format algorithm. It relies on the notion that the spherical waves radiated by the sensor are approximately planar over the scene to be imaged. This amounts to a narrow-beam restriction, and many fine resolution SAR applications fall into this category. Wide-beam collections are encountered when using low frequencies, as in foliage penetration, or when imaging at short ranges. The latter is also known as near-field imaging. These situations call for image reconstruction techniques such as the ω - k algorithm [4,17] or backprojection that account for the spherical nature of the propagating wavefronts. Backprojection is the most general form of image reconstruction, but it is also very expensive compared with PFA.

The nature and impact of linear, quadratic, sinusoidal, and random phase errors were discussed. Phase error effects tend to fall into the following categories: (1) loss of resolution

due to broadening of the PSR, (2) image artifacts such as multiple copies of a single object, and (3) a rise in the background noise. Two important methods for autofocus, or post-processing image enhancement, were presented. Autofocus is a standard component of many SAR image formation systems because it helps remove any residual phase errors that are otherwise unaccounted for. A primary example is uncompensated platform motion.

The SAR system parameters can be used to predict image contrast, which is a good measure of image quality. This analysis is often used in the design stage to ensure that the entire hardware and software chain can produce imagery that meets the needs of the end users. Besides thermal noise, major factors affecting image contrast are range and cross-range ambiguities, quantization noise, and the character of the point scatterer response in the image itself. Image rating scales are also used to assess image quality. These are more subjective, as they rate the image in terms of its ability to provide information for specific tasks. A radar capable of making fine resolution high-contrast imagery can still produce poor imagery. For example, a poorly chosen collection geometry might cause a town to be obscured by shadows from nearby mountains.

Readers interested in exploring spotlight SAR in more detail are encouraged to review the books by Curlander and McDonough [3], Jakowatz et al. [4], and Carrara et al. [5]. Together, these three volumes provide a thorough reference library for SAR system design and image formation. Curlander and McDonough [3] offer a comprehensive view of radar system engineering and explain many ideas from the standpoint of basic physics. It is oriented toward stripmap imaging, so the discussion of image reconstruction algorithms is of limited use. Nevertheless, the authors provide many details of SAR system design that are difficult to find elsewhere in the literature. The books by Jakowatz et al. [5] and Carrara et al. [4] offer a modern view of image formation. Carrara et al. is probably the most comprehensive reference on the subject of spotlight SAR up to and including image formation. It is particularly valuable because it covers many real-world effects such as allowable deviations from an ideal flight path and correcting PFA imagery for wavefront curvature effects. Jakowatz et al. focus on the polar format algorithm and the projection slice theorem, and these ideas are used to develop the procedures for change detection and interferometry. Their discussion is particularly clear and accessible.

A more recent reference is the February 2010 issue of *IEEE Transactions on Geoscience and Remote Sensing* [35]. It is dedicated to the German TerraSAR-X satellite mission, and this issue contains articles on many interesting topics such as system design, calibration procedures, and advanced processing modes such as moving target detection and sea current observation. There are many other books on the subject of SAR and its more advanced applications, some of which are discussed in Chapter 8. Although the basics of image formation are largely settled, the SAR literature is constantly growing. Cutting-edge research is being performed by institutions in several countries, so it is always wise to follow the current literature for innovations and trends.

6.10 | REFERENCES

- [1] A. W. Love. In memory of Carl A. Wiley. *IEEE Antennas and Propagation Society Newsletter*, pages 17–18, June 1985.
- [2] Carl A. Wiley. The clipper ships of space. *Astounding Science Fiction*, 1951.
- [3] John C. Curlander and Robert N. McDonough. *Synthetic Aperture Radar: Systems and Signal Processing*. John Wiley and Sons, Inc., 1991.

- [4] Walter G. Carrara, Ron S. Goodman, and Ronald M. Majewski. *Spotlight Synthetic Aperture Radar Signal Processing Algorithms*. Artech House, 1995.
- [5] Charles V. Jakowatz, Daniel E. Wahl, Paul H. Eichel, Dennis C. Ghiglia, and Paul A. Thompson. *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. Kluwer Academic Publishers, 1996.
- [6] A. V. Oppenheim and R. W. Schaffer. *Discrete-Time Signal Processing*. Prentice Hall, 2009.
- [7] David C. Munson, Jr., James Dennis O'Brien, and W. Kenneth Jenkins. A tomographic formulation of spotlight-mode synthetic aperture radar. *Proceedings of the IEEE*, 71(8):917–925, August 1983.
- [8] David C. Munson, Jr. and Robert L. Visentin. A signal processing view of strip-mapping synthetic aperture radar. *IEEE Transactions of Acoustics, Speech, and Signal Processing*, 37(12):2131–2147, December 1989.
- [9] R. N. Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill Inc., 1965.
- [10] Arch W. Naylor and George R. Sell. *Linear Operator Theory in Engineering and Science*. Springer-Verlag, 1982.
- [11] T. K. Moon and W. C. Stirling. *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 1999.
- [12] Fredric J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1), January 1978.
- [13] R. Haberman. *Applied Partial Differential Equations with Fourier Series and Boundary Value Problems*. Pearson Prentice Hall, 2004.
- [14] Mark A. Richards, James A. Scheer, and William A. Holm, editors. *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Inc., 2010.
- [15] C. H. Casteel, L. A. Gorham, M. J. Minardi, S. Scarborough, and K. D. Naidu. A challenge problem for 2D/3D imaging of targets from a volumetric data set in an urban environment. *Proceedings of the SPIE 6568*, April 2007.
- [16] D. A. Cook and D. P. Campbell. Synthetic aperture imaging via backpropagation using commodity graphics processors. In *Proceedings of the International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*. Institute of Acoustics, 2010.
- [17] David W. Hawkins. *Synthetic Aperture Imaging Algorithms: with application to wide bandwidth sonar*. PhD thesis, Department of Electrical and Electronic Engineering, University of Canterbury, Christchurch, New Zealand, 1996.
- [18] J. G. Mehliis. Synthetic aperture radar range-azimuth ambiguity design and constraints. *IEEE International Radar Conference*, pages 143–152, 1980.
- [19] C. Elachi. *Spaceborne Radar Remote Sensing*. IEEE Press, 1987.
- [20] S. Musman, D. Kerr, and C. Bachman. Automatic recognition of isar ship images. *IEEE Transactions on Aerospace and Electronic Systems*, 32(4), October 1996.
- [21] Armin W. Doerry. Ship dynamics for maritime ISAR imaging. Technical report, Sandia National Laboratories, 2008.
- [22] David C. Munson, Jr. and Jorge L. C. Sanz. Image reconstruction from frequency-offset fourier data. *Proceedings of the IEEE*, 72(6):661–669, June 1984.
- [23] D. A. Cook and D. C. Brown. Analysis of shadow contrast for synthetic aperture sonar imagery. In *Proceedings of the International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*. Institute of Acoustics, 2010.

- [24] Mehrdad Soumekh. *Fourier Array Imaging*. Prentice-Hall, Inc., 1994.
- [25] Daniel A. Cook and Daniel C. Brown. Analysis of phase error effects on stripmap SAS. *IEEE Journal of Oceanic Engineering*, 34(3):250–261, July 2009.
- [26] B. Widrow. Statistical theory of quantization. *IEEE Transactions on Instrumentation and Measurement*, 45(2):353–361, April 1996.
- [27] H. Zumbahlen, editor. *Linear Circuit Design Handbook*. Newnes, Norwood, MA, 2008.
- [28] Federation of American Scientists: National Image Interpretability Rating Scales <http://www.fas.org/irp/imint/niirs.htm>.
- [29] Charles E. Cook and Marvin Bernfeld. *Radar Signals: An Introduction to Theory and Application*. Artech, 1993.
- [30] Max Born and Emil Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference, and Diffraction of Light*. Cambridge University Press, 1999.
- [31] Peter T. Gough and David W. Hawkins. Imaging algorithms for a strip-map synthetic aperture sonar: Minimizing the effects of aperture errors and aperture undersampling. *IEEE Journal of Oceanic Engineering*, 22(1):27–39, January 1997.
- [32] G. Fornaro. Trajectory deviations in airborne SAR: Analysis and compensation. *IEEE Transactions on Aerospace and Electronic Systems*, 35(3):997–1009, July 1999.
- [33] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1984.
- [34] Mark A. Richards. Coherent integration loss due to white gaussian phase noise. *IEEE Signal Processing Letters*, 10(7):208–210, July 2003.
- [35] Special Issue on TerraSAR-X: Mission, Calibration, and First Results. *IEEE Transactions of Geoscience and Remote Sensing*, 48(2), February 2010.

6.11 | PROBLEMS

1. Explain the qualitative difference between SAR and optical imagery at low grazing angles and at high grazing angles.
2. Where is the radar in any given image? What clues can you use to decide?
3. What is the relationship between range resolution in the slant plane and the ground plane? What about cross-range resolution?
4. Derive the angular and along-track sampling requirements, $\delta\phi$ and δx , for a collection where the radar is squinted by θ_{sq} relative to broadside in the ground plane (or equivalently by θ_{cone} relative to the velocity vector).
5. Use a computer programming language to write a truncated sinc interpolator. Use the software to reproduce the results shown in Figure 6-16.
6. Describe the major sources of noise in SAR imagery. Explain the difference between additive and multiplicative noise. To which category does receiver phase noise belong? What about rainfall within the scene?
7. Azimuth ambiguities are usually treated as uncorrelated noise when predicting image quality. However, very bright features such as buildings, towers, and bridges can appear in a SAR image as azimuth ambiguities. Will these ambiguous features be properly focused by the polar format algorithm?

8. The transmitted bandwidth of a radar is often idealized as having a constant value across the band. What happens to the range profile if the spectrum amplitude exhibits a sinusoidal variation (often called passband ripple)? Is the cross-range character of the image affected?
9. Figures 6-11 and 6-12 depict the fact that spotlight SAR imagery is usually formed by retaining the largest rectangular region that can be inscribed in the annular spectral support of the radar data. Write the expressions for the dimensions of this rectangle.
10. Assuming the cross-range resolution is fixed, does increasing the radar's center frequency reduce or increase the effect of quadratic phase error?

Stripmap SAR

Gregory A. Showman

Chapter Outline

7.1	Introduction	259
7.2	Review of Radar Imaging Concepts	264
7.3	Doppler Beam Sharpening Extensions	271
7.4	Range-Doppler Algorithms	286
7.5	Range Migration Algorithm	305
7.6	Operational Considerations	318
7.7	Applications	327
7.8	Summary	330
7.9	Further Reading	331
7.10	References	332
7.11	Problems	333

7.1 INTRODUCTION

Synthetic aperture radar (SAR) is a combination of radar hardware, waveforms, signal processing, and relative motion that strives to render radar returns from stationary targets and scenes in a product reminiscent of an overhead optical photograph. The principal product of any basic SAR implementation is a fine-resolution two-dimensional intensity image of the illuminated area. Resolution, the ability to distinguish between closely spaced point targets, is provided for in the range dimension by waveform bandwidth and pulse compression techniques. Resolution in the orthogonal, cross-range dimension is afforded by the known motion of the radar platform with respect to the imaged scene and associated matched filters and data transformations.

To pique the curiosity in readers new to the specialty, Figure 7-1 shows two pictures of a portion of the Albuquerque Airport: one a traditional optical photograph taken from overhead and the other a SAR product. At first glance the image properties appear to be similar, but a closer examination reveals features peculiar to SAR. For example, the SAR image shows an augmentation of man-made objects like buildings, which cast shadows down-range with respect to the radar location. Radar images have the additional advantage that they can be acquired at night and under all weather conditions. Furthermore, low frequency operation allows radar to penetrate foliage or the ground and create images of

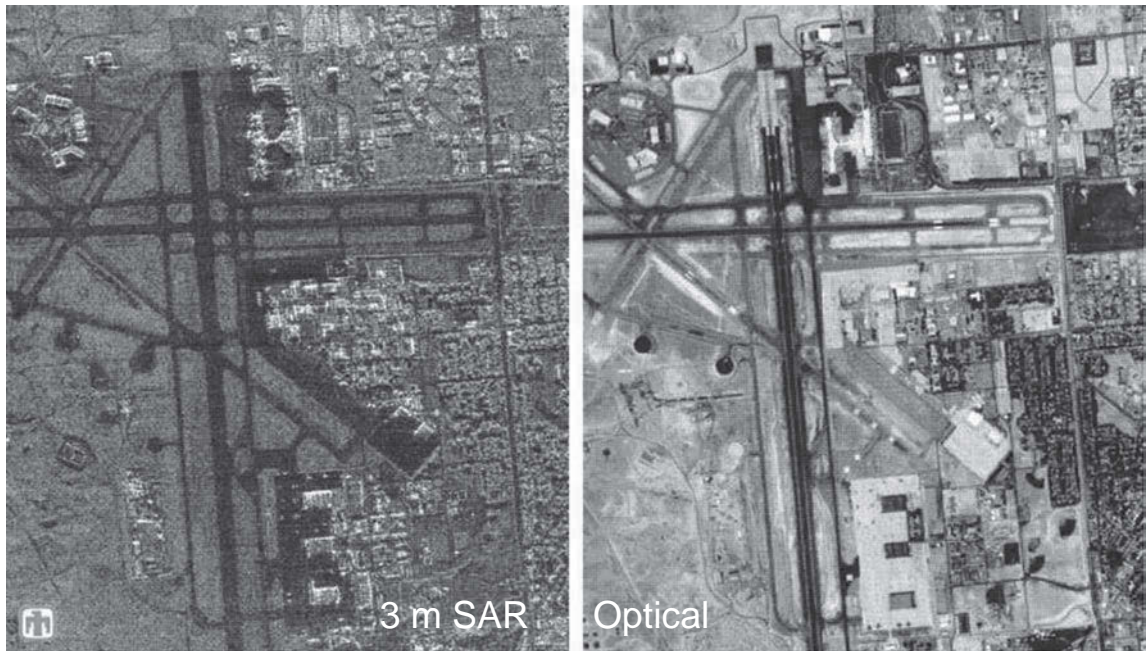


FIGURE 7-1 ■ Two renderings of Albuquerque Airport with SAR on the left and an overhead optical photograph on the right. (Images courtesy of Sandia National Laboratories.)

hidden and buried objects. Finally, phase information in complex SAR imagery may be exploited to estimate terrain height or changes to the scene over time.

A great variety of platform motion histories are capable of producing usable SAR imagery. One specific case is known as stripmap SAR. In SAR vernacular, *stripmap* may refer to the data acquisition geometry, a corresponding image formation technique, or both. The key characteristics of a stripmap collection are:

- The platform moves in a straight line. (This is in contrast to the curved and circular paths often employed by spotlight systems.)
- Transmit and receive antennas are pointed to a fixed angle, frequently broadside, normal to the direction of flight.
- Transmit and receive beam patterns are not steered in angle during data acquisition (as opposed to spotlight mode).

Stripmap image formation techniques are designed to produce focused SAR images consistent with these conditions.

While the stripmap geometry may appear overly specific and constraining, there is a rich history of collecting and processing SAR data in such a manner, particularly within the remote sensing community. The SAR image in Figure 7-1 is an excellent example of a stripmap product with respect to both the collection geometry and image formation. The following features motivate the stripmap mode:

- Stripmap yields the finest cross-range resolution possible without sparse or interrupted coverage.

- Stripmap operation requires no mechanical or electronic beam-steering of the radar antenna.¹
- The straight-line flight path intrinsic to stripmap serves to maximize the area surveyed. An arbitrarily large region may be imaged in stripmap by an airborne platform by collecting along a series of parallel straight lines in a raster-like pattern.
- Stripmap is a natural collection mode for satellites, whose orbits appear, to first order, as linear paths with respect to a point on the ground.

For these reasons, stripmap is a SAR subtopic worthy of a detailed exposition. In addition, many well-known SAR image formation algorithms are tailored to the stripmap collection geometry. Finally, the stripmap framework offers an interesting contrast to the spotlight SAR model presented elsewhere in this volume.

7.1.1 Organization

We begin with a review of the salient developments and results presented in [1]. Particular attention is paid to the point spread response (PSR), which is used to derive the very simple Doppler beam sharpening (DBS) image former. DBS is an acceptable technique for coarse resolutions and small scene sizes. Under more challenging conditions DBS struggles with nonlinear phase histories and migration of scatterers through range bins. These effects may be mitigated to a limited degree with azimuth dechirp (AD) and range migration compensation (RMC) stages. A more holistic approach is to use the PSR explicitly to build matched filters to the expected scatterer histories, which leads directly to the range-Doppler algorithm (RDA) family of image formers. Theoretically, RDA methods are limited in performance to a small down-range extent due to the variation in the PSR form with scatterer range. The range migration algorithm (RMA) begins with RDA but adds a Stolt interpolation stage to overcome limitations in down-range extent. Next, stripmap implementation topics are covered, such as waveform selection and area coverage rate calculations. Finally, common stripmap applications are described, including remote sensing and foliage penetration (FOPEN) imaging.

7.1.2 Key Points

- Stripmap SAR employs a linear collection geometry and a broadside-looking antenna.
- This mode of operation achieves the finest possible cross-range resolution while surveying the passing terrain without gaps in coverage.
- The PSR, the manifestation of a point target in the raw data, takes the form of a hyperbola.
- The family of simple DBS image formation techniques may be formulated using the hyperbolic PSR.
- More sophisticated image formers, like RDA and RMA, have their basis in matched filter theory and derive directly from the PSR.

¹ While the ideal mode does not involve steering the transmit or receive beams, airborne stripmap systems usually require some limited steering to compensate for platform roll, pitch, and yaw over the collection, thereby ensuring the commanded beam-pointing angle is maintained.

- RMA is the preferred technique for stripmap collections, as it is computationally efficient and makes no approximations to the PSR.
- Stripmap is a common mode for remote sensing and FOPEN SAR systems.

7.1.3 Notation

Many of the variables in this chapter are as follows:

c = velocity of propagation (speed of light, m/s)

v = platform ground speed (m/s)

f = radio frequency (RF) (Hz)

$\omega = 2\pi f$ = radian RF (rad/s)

λ = RF wavelength (m)

f_c = center frequency (Hz)

$\omega_c = 2\pi f_c$ = radian center RF (rad/s)

λ_c = center wavelength (m)

PRF = pulse repetition frequency (Hz)

B_r = waveform bandwidth (Hz)

B_d = Doppler bandwidth (Hz)

B_{2R} = Doppler bandwidth over null-to-null beamwidth (Hz)

B_{SAR} = Doppler bandwidth applied to SAR processing (Hz)

θ_{int} = integration angle (radians)

θ_g = grazing angle (radians)

θ_B = 3 dB azimuth beamwidth (radians)

θ_R = azimuth peak-to-null beamwidth (radians)

ϕ_R = elevation peak-to-null beamwidth (radians)

k_u = along-track spatial frequency (radians/m)

$k_{u,Mig}$ = spatial frequency due to range migration (radians/m)

k_r = RF spatial frequency; related to ω (radians/m)

k_{r0} = RF spatial frequency at center of band (radians/m)

k_x = along-track spatial frequency in SAR image (radians/m)

k_y = cross-track spatial frequency in SAR image (radians/m)

d = along-track sampling interval (m)

D = antenna length (m)

D_{az} = antenna horizontal length (m)

D_{el} = antenna vertical height (m)

D_{SAR} = SAR along-track coherent collection/processing distance (m)

K_{AZ} = relaxation (decrease) in along-track (lower) PRF limit, 1 to 2 (unitless)

K_{EL} = relaxation (increase) in range-ambiguity (upper) PRF limit, 1 to 2 (unitless)

ΔR = downrange resolution (m)

ΔCR = crossrange resolution (m)

$\Delta CR(stripmap)$ = limit on achievable stripmap crossrange resolution (m)

DOF = depth of focus (m)

- u = along-track location of platform (m)
 x = along-track (crossrange) location of scatterer (m)
 r = cross-track (downrange) location of scatterer (m)
 r_0 = reference cross-track (downrange) to imaged swath (m)
 Δr = differential cross-track (downrange) with respect to reference (m)
 R = slant-range from platform to scatterer (m)
 t = time-delay into data (s)
 x_S = crossrange half-extent of scene (m)
 y = cross-track (downrange) location of pixel in image (m)
- $g(x, r)$ = reflectivity of scene (complex RCS, m)
 $d(u, t)$ = recorded data (complex RCS, m)
 $d_{Dechirp}(u, t)$ = data after azimuth dechirp (complex RCS, m)
 $d_{RMC}(u, t)$ = data after range migration compensation (complex RCS, m)
 $D(k_u, t)$ = along-track Fourier transform of data (complex RCS, m)
 $D(k_u, \omega)$ = 2-D Fourier transform of data (complex RCS, m)
 $D_{RMC}(k_u, \omega)$ = 2-D Fourier transform of data after RMC (complex RCS, m)
 $f(x, r)$ = SAR image (complex RCS, m)
- $\delta_D(t)$ = Dirac delta function
 $h(u, t; x, r)$ = PSR (complex RCS, m)
 $h(u, t; r)$ = reference PSR (complex RCS, m)
 $h(u, t; r_0)$ = reference PSR at r_0 (complex RCS, m)
- ACR = area coverage rate (m^2/s)
 T_{Swath} = pulse time over swath (s)
 L_{Swath} = length of swath on the ground (m)
 $L_{Swath, BW}$ = length of elevation-beamwidth-limited L_{Swath} (m)
 U_{SAR} = SAR system utility (m/s)
- Δt = time-delay progression due to platform motion (s)
 t_0 = reference time delay for reference range r_0 (s)
 t_{Shift} = time shift for support of PSR in (k_u, t) (s)
 t_{Dif} = differential time shift with respect to reference PSR (s)
 $\Delta\psi$ = phase progression due to platform motion (radians)
 $\Delta\psi_{QT}$ = quadratic term in phase progression (radians)
 ψ_{Mig} = phase progression due to range migration (radians)
 ψ_{PSR} = 2-D frequency phase function for PSR (radians)
 ψ'_{PSR} = 1st approximation to phase function for PSR (radians)
 ψ''_{PSR} = 2nd approximation to phase function for PSR (radians)
- $d(t)$ = generic 1-D data record (complex voltage)
 $h(t)$ = generic 1-D filter impulse response (complex voltage)
 $h_0(t)$ = generic 1-D matched filter impulse response (complex voltage)
 $s(t)$ = generic 1-D signal (complex voltage)
 $y(t)$ = generic 1-D filter output (complex voltage)
 $D(\omega)$ = generic 1-D data frequency response (complex voltage)
 $H(\omega)$ = generic 1-D filter frequency response (complex voltage)

$H_0(\omega)$ = generic 1-D matched filter frequency response (complex voltage)
 $S(\omega)$ = generic 1-D signal frequency response (complex voltage)
 $Y(\omega)$ = generic 1-D filter frequency output (complex voltage)
 $d(u, t)$ = generic 2-D data record (complex voltage)
 $h(u, t)$ = generic 2-D filter impulse response (complex voltage)
 $h_0(u, t)$ = generic 2-D matched filter impulse response (complex voltage)
 $s(u, t)$ = generic 2-D signal (complex voltage)
 $y(u, t)$ = generic 2-D filter output (complex voltage)
 $D(k_u, \omega)$ = generic 2-D data frequency response (complex voltage)
 $H(k_u, \omega)$ = generic 2-D filter frequency response (complex voltage)
 $H_0(k_u, \omega)$ = generic 2-D matched filter frequency response (complex voltage)
 $S(k_u, \omega)$ = generic 2-D signal frequency response (complex voltage)
 $Y(k_u, \omega)$ = generic 2-D filter frequency output (complex voltage)

7.1.4 Acronyms

Commonly used acronyms used in this chapter include the following:

ACR	area coverage rate
AD	azimuth dechirp
AWGN	additive white Gaussian noise
CPI	coherent processing interval
CRP	center reference point
CSA	chirp scaling algorithm
DBS	Doppler beam sharpening
DOF	depth of focus
FFT	fast Fourier transform
FOPEN	foliage penetration
LFM	linear frequency modulation
LOS	line of sight
MCL	motion compensation to a line
MCP	motion compensation to a point
PSR	point spread response
RDA	range-Doppler algorithm
RMA	range migration algorithm
RMC	range migration compensation
SAR	synthetic aperture radar
SAS	synthetic aperture sonar
SRC	secondary range compression
SNR	signal-to-noise ratio

7.2 | REVIEW OF RADAR IMAGING CONCEPTS

[1] emphasizes the signal processing paradigm to explain SAR and construct image formation algorithms. Key results from that chapter are summarized in this section.

7.2.1 Resolution and Sampling

A key principle of the signal processing mindset is that resolution of closely spaced targets and, more generally, information about a scene is garnered by collecting returns over as wide a frequency and angle extent as possible.

Frequency content comes directly from the waveform bandwidth. Down-range resolution, ΔR , is given by

$$\Delta R = \frac{c}{2B_r} \quad (7.1)$$

where c is the propagation speed of the imaging energy (the speed-of-light for SAR), and B_r is the radio frequency (RF) bandwidth of the radar waveform. For example, for the speed-of-light equal to approximately 3×10^8 m/s and given a waveform bandwidth of 1 GHz, the down-range resolution is about 0.15 m, around 6 inches. In general, *down-range* corresponds to the direction along the line of sight (LOS) between the radar and the imaged scene. For stripmap SAR, *down-range* specifically corresponds to a direction normal to the flight path. Because this path is referenced with respect to the ground-track, the projection of the flight path vertically onto the ground, this dimension is often referred to as *cross-track*.

Angle diversity is generated by the motion of the platform with respect to the imaged scene. Translational motion of the platform perpendicular to the LOS to the scene causes a change of aspect angle to the radar from the point of view of a target in the scene. The total aspect angle change over a coherent collection time is known as the integration angle θ_{int} . The cross-range resolution, ΔCR , is given by

$$\Delta CR = \frac{\lambda_c}{2\theta_{int}} \quad (7.2)$$

where λ_c is the wavelength at the center (carrier) frequency. For example, at an RF of 10 GHz (0.03 m wavelength), and integration angle of 6° (approximately 1/10 of a radian), the cross-range resolution is about 0.15 m. In general, *cross-range* corresponds to a direction normal to the LOS and horizontal to the terrain. For stripmap SAR, *cross-range* specifically corresponds to a direction along the platform ground-track; hence, this dimension is often denoted *along-track*.

Integration angle against a scene may be achieved through a variety of platform paths. Three examples—stripmap, linear spotlight, and circular spotlight—are shown in Figure 7-2. Achievable integration angle increases and, hence, cross-range resolution improves from left to right. This chapter focuses on the stripmap case, as the title suggests, though we shall see that stripmap results may be applied to the linear spotlight case as well.

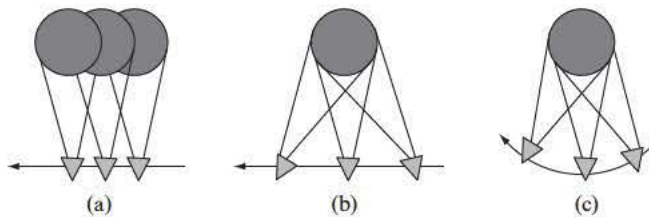


FIGURE 7-2 ■ Three collection options for achieving integration angle. (a) Stripmap. (b) Spotlight with mechanical or electronic steering of the real antenna beam onto a scene of interest. (c) Spotlight realized by flying a circle about the scene of interest.

For stripmap collections it was found that the achievable integration angle is limited to no more than the 3 dB beamwidth of the physical radar antenna. Applying a common antenna rule-of-thumb—that the 3 dB beamwidth is equal to wavelength divided by aperture size—to (7.2), a constraint on stripmap cross-range resolution was established as

$$\Delta CR(\text{stripmap}) \geq \frac{D}{2} \quad (7.3)$$

where D is the horizontal length of the real antenna. For example, an antenna 1 m in size can produce a stripmap SAR image with a cross-range resolution no better than 0.5 m.

To record data unambiguously, information along-track must be sampled at a sufficiently high spatial rate. The maximum interval between samples d is set by the null-to-null beamwidth of the real antenna. Using another rule-of-thumb—that the peak-to-null extent is equal to wavelength divided by aperture size—it was found that the interval between samples must be

$$d \leq \frac{D}{4} \quad (7.4)$$

For example, an antenna 1 m in size requires an along-track sampling interval no larger than 0.25 meters. If the platform is moving at 100 m/s this requirement is met by a pulse repetition frequency (PRF) of 400 Hz or higher.

7.2.2 Point Spread Response

A second tenet from the signal processing paradigm is that high-fidelity image formation techniques may be developed through knowledge of the point spread response (PSR), the manifestation of a point target in the measured data.

Figure 7-3 shows domains and transformations over the full end-to-end SAR procedure. The top diagram depicts the data acquisition geometry. Cross-track (down-range) and along-track (cross-range) dimensions are denoted by r and x , respectively. In this development r is distance in the slant plane (along the line of sight from the platform to the ground); later in the chapter we discuss mappings from down-range in the slant plane to downrange in the ground plane. The complex scene content is represented by the function $g(x, r)$. The diagram at left depicts the recorded data as a function of fast-time delay t and along-track position of the platform u . The complex data content is represented by the function $d(u, t)$. The last diagram depicts the final SAR image, which is hopefully an accurate rendering of the original scene. Hence, it shares the same dimensions, x and r . The complex image content is represented by the function $f(x, r)$.

The slant-range R from the platform to a location in the scene is calculated by referring to Figure 7-4 and applying the Pythagorean theorem:

$$R(u; x, r) = \sqrt{(u - x)^2 + r^2} \quad (7.5)$$

This function takes the form of a hyperbola. The shape of the PSR is the same for all scatterers at the same cross-track location (constant r) and different along-track locations. However, the shape changes as cross-track varies, so the PSR for scatterers near the ground-track are severely bowed whereas the PSR for scatterers farther away is flatter.

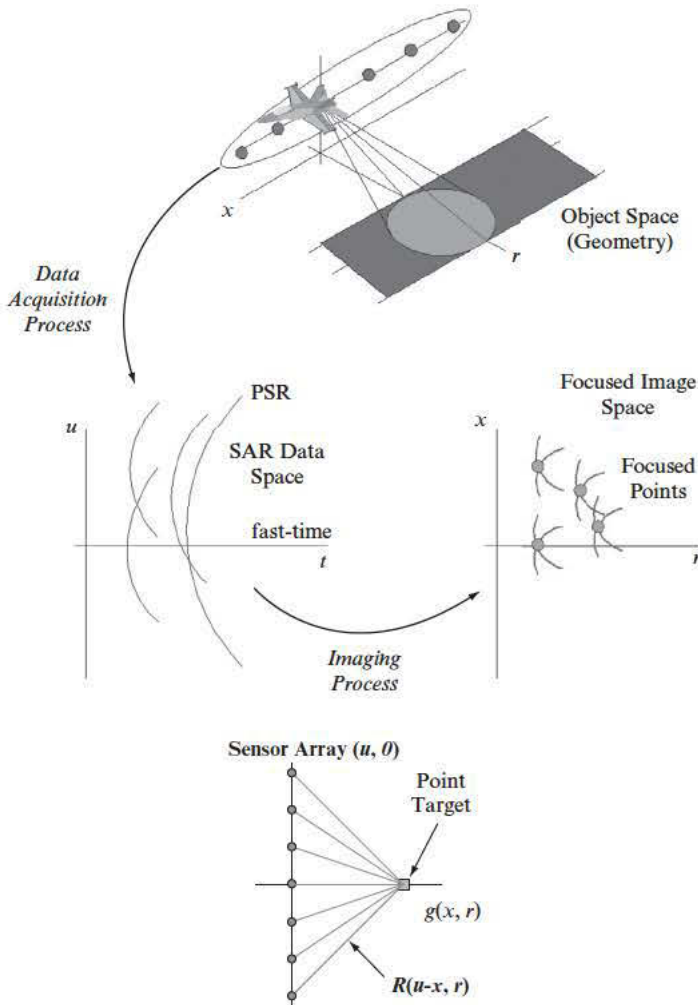


FIGURE 7-3 ■ The SAR process involves three domains and two transformations. Data acquisition provides the transformation from scene reflectivity to raw data, while image formation operates on the raw data to yield the final SAR image.

FIGURE 7-4 ■ Variation in slant-range R to scatterer at (x, r) due to platform motion.

The PSR maps to time delay into the raw data as

$$t(u; x, r) = \frac{2}{c} R(u; x, r) = \frac{2}{c} \sqrt{(u - x)^2 + r^2} \quad (7.6)$$

A high-fidelity image formation technique called backprojection was developed in the chapter. This algorithm operates by directly exploiting the PSR to build a separate filter matched to each and every image pixel. We needed to represent the PSR mathematically to justify backprojection, and we found the PSR takes the form of a Dirac-delta function, $\delta_D(t)$. The Dirac-delta function has a value of zero for all t except at $t = 0$, where the function takes on a value of infinity, and has a total area (integral over all time t) equal to 1. The PSR used to build an imaging reference function h (e.g., as a matched filter for backprojection) is

$$\begin{aligned} h(u, t; x, r) &= \delta_D(t - t(u; x, r)) \\ &= \delta_D\left(t - \frac{2}{c} \sqrt{(u - x)^2 + r^2}\right) \end{aligned} \quad (7.7)$$

TABLE 7-1 ■ Survey of SAR Image Formation Algorithms

Category	Features	Limitations	Algorithms	Notes
Doppler	Requires only a 1-D fast Fourier transform (FFT) after pulse compression	No accounting for migration of scatterers through range bins or phase orders above linear	Doppler beam sharpening (DBS)	Fourier transform provides the matched filter
			DBS with azimuth dechirp	Accounts for quadratic phase error (QPE) by applying a conjugate quadratic modulation to the data prior to DBS
Fourier matched filter	Match filtering with the point spread response (PSR) implemented efficiently by multiplication in the frequency domain	Data acquisition must be linear with uniform spacing between along-track samples	Range-Doppler algorithm	Assumes a single PSR is good for the entire scene, so focused swath depth is limited
			Range migration algorithm (RMA)	Range Doppler algorithm combined with Stolt data interpolation provides error-free imaging over deep swaths
			Chirp scaling algorithm	Exploits direct sampling of linear frequency modulated (LFM) waveforms to realize a computationally efficient approximation to the Stolt interpolation
			Range stacking algorithm	Like RMA, but less computationally efficient
Tomographic	Projection-slice view of data acquisition; efficient image formation follows	Scene size; data must be time delayed and phase corrected prior to image formation	Rectangular formatting algorithm	Analogous to DBS in simplicity; capable of fine resolution, but scene size severely restricted
			Polar formatting algorithm	2-D data interpolation provides quality imaging at fine resolutions over reasonable scene sizes; scene extents ultimately limited by geometric considerations
Inverse	A more holistic, mathematical view of data acquisition and image formation	Demanding processing requirements, and sometimes memory requirements as well	Matched filtering	Employs a product of the data and a reference response; implementable in either time-distance or frequency domains
			Constant aperture backprojection	Delay and sum imaging; same data records used for all pixels
			Constant integration angle backprojection	Like constant aperture backprojection, but different subset of the overall data applied to each pixel to realize a constant integration angle
			Filtered backprojection	Wideband/wide-angle data are preequalized to improve impulse response in the final image
			Fast backprojection	Family of algorithms that garner computational efficiencies by making minor approximations to basic backprojection
			Quad-tree	Fast backprojection using multiresolution topologies
			Constrained inversion	SAR imaging treated as a mathematical inverse problem

The PSR, then, is defined in the raw data domain (u, t) , is a function of scatterer location (x, r) , and is zero everywhere except at times equal to the time delay to the scatterer return.

7.2.3 Image Formation Survey

Dozens of different image formation techniques have been developed and implemented over the last half-century. These many algorithms were grouped into four categories in [1]; the table summarizing this taxonomy is reproduced in Table 7-1. Backprojection, the most popular of the inverse methods, was discussed briefly in [1]. Tomographic methods are geared toward spotlight collections, so these are covered in detail in the companion spotlight chapter. The Doppler and Fourier matched filter categories are tailored to stripmap acquisitions, so these algorithms are emphasized in this chapter as low- and high-fidelity options, respectively.

7.2.4 Doppler Beam Sharpening

The oldest and most fundamental image formation technique is DBS, which operates by exploiting a one-to-one mapping between Doppler frequency and azimuth angle. Referring to Figure 7-5, the circles of slant-range bins and angles of Doppler filter responses compose a range-Doppler map. For a small scene size at a long range from the radar, the map approximates a rectilinear down-range–cross-range image over the mainbeam.

DBS may be derived using the PSR. Starting with the time-delay PSR and assuming the along-track difference between the platform and scatterer locations is much less than the down-range distance to the scatterer, that is, $(u - x) \ll r$, we may apply the binomial approximation

$$\begin{aligned} t(u; x, r) &= \frac{2}{c} \sqrt{(u - x)^2 + r^2} \\ &\approx \frac{2}{c} \left(r + \frac{u^2}{2r} - u \frac{x}{r} + \frac{x^2}{2r} \right) \end{aligned} \quad (7.8)$$

We are concerned with how a scatterer location is encoded into platform along-track u . The first and last terms do not depend on platform location, so these terms are omitted

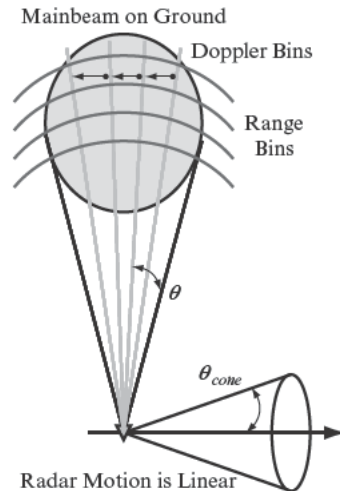


FIGURE 7-5 ■ Arranging ground returns as a function of slant-range and Doppler yields a 2-D rendering of the scene of interest.

to form a time-delay variation, Δt ,

$$\Delta t(u; x, r) \approx \frac{2}{c} \left(\frac{u^2}{2r} - u \frac{x}{r} \right) \quad (7.9)$$

The time-delay variation generates a phase progression in the raw data $\Delta\psi$ with respect to the carrier frequency, f_c ,

$$\begin{aligned} \Delta\psi &= 2\pi f_c \Delta t(u; x, r) \\ &\approx \frac{4\pi}{\lambda_c} \left(\frac{u^2}{2r} - u \frac{x}{r} \right) \end{aligned} \quad (7.10)$$

For short SAR collections (small u) the quadratic term may be ignored, leaving

$$\Delta\psi \approx -\frac{4\pi}{\lambda_c} u \left(\frac{x}{r} \right) \quad (7.11)$$

The along-track location of the scatterer x is encoded onto the data as a linear phase progression with platform location u . Taking the derivative of the phase progression with respect to platform along-track location yields a spatial frequency, k_u ,

$$\begin{aligned} k_u &\equiv \frac{\partial}{\partial u} \Delta\psi(u) \\ &\approx -\frac{4\pi}{\lambda_c} \left(\frac{x}{r} \right) \end{aligned} \quad (7.12)$$

The spatial frequency in radians/meter is like a Doppler frequency in Hz but is tied to the location, and not time, at which a sample was recorded.

The spatial frequency is constant over the dwell for a given scatterer location, and the Fourier transform is a known matched filter for constant frequency signals. Therefore, DBS begins by Fourier transforming the recorded data from u to k_u

$$D(k_u, t) = \int_{-\infty}^{\infty} d(u, t) e^{-jk_u u} du \quad (7.13)$$

and then mapping spatial frequency and time delay to cross-range and down-range

$$f(x, r) = D\left(k_u \left(-\frac{\lambda_c r_0}{4\pi}\right) \rightarrow x, t \left(\frac{c}{2}\right) \rightarrow r\right) \quad (7.14)$$

where r_0 is the cross-track distance to scene center. A flowchart for DBS is given in Figure 7-6.

DBS was demonstrated in [1] and performed well on a simulated data set having the following characteristics:

- Carrier frequency = 9.5 GHz
- RF bandwidth = 3 MHz
- Integration angle = 0.02°
- Resolution = 50 m (both dimensions)
- Range to scene center = 50 km
- Scene size = 4 km by 4 km

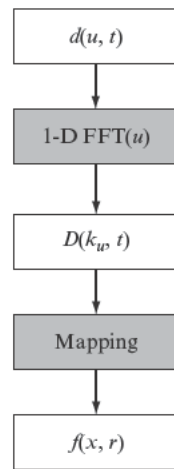


FIGURE 7-6 ■ DBS processing performs a spatial Fourier transform over the pulse history followed by mapping the data into scene coordinates.

7.3 | DOPPLER BEAM SHARPENING EXTENSIONS

Supplemental processing stages are developed for DBS to allow the image formation technique to function over larger scene sizes and finer resolutions.

7.3.1 Azimuth Dechirp

Producing imagery with finer resolution requires separate actions in each dimension. Increasing the waveform bandwidth is the means to improving down-range resolution. More bandwidth is realized by raising the chirp rate on an LFM pulse, for example. Increasing the integration angle is the means to improving cross-range resolution. If down-range to the scene is fixed, this is achievable only by adding to the coherent collection time and distance.

Figure 7-7 depicts this improvement in resolution as compared to Figure 7-5. Smaller down-range and cross-range bins are evident. However, the platform must move some

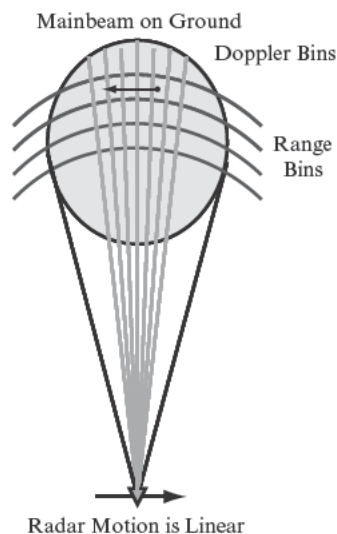
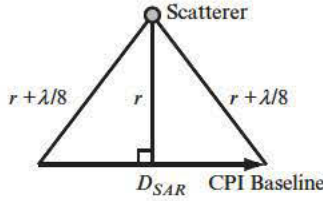


FIGURE 7-7 ■ SAR collection with increasingly fine resolution and consequent scatterer drift over cross-range resolutions cells.

FIGURE 7-8 ■
Geometry limiting
slant-range change
to a broadside
scatterer to
one-eighth of a
wavelength.



distance (through some integration angle) to achieve this cross-range resolution. Equivalently, we can imagine that the platform is stationary and that the scatterers move past the radar. With increasingly finer resolution there will come a point where scatterers drift across cross-range bins during the collection time, resulting in cross-range smearing in the final DBS image. Such a scatterer is represented by the black arrow in the mainbeam in Figure 7-7.

From the Doppler point of view, the Doppler shift to the scatterer is changing over the dwell. One can show that this change in Doppler is negligible if the slant-range variation to a target having zero Doppler frequency (at broadside) is less than one-eighth of a wavelength. (This constraint corresponds to a two-way slant-range limit of one-fourth of a wavelength, a phase variation of 90° .) This limit on geometry is shown in Figure 7-8 for a total along-track collection distance of D_{SAR} .

Again reverting to the Pythagorean theorem and assuming the collection interval is much smaller than the range to the scene, $D_{SAR} \ll r$, a limit on the coherent processing interval (CPI) baseline is easily found as

$$D_{SAR} \leq \sqrt{r\lambda_c} \quad (7.15)$$

Referring to the expression for cross-range resolution in (7.2) and noting that the integration angle is approximately equal to the ratio of r and D_{SAR} yields

$$\Delta CR = \frac{\lambda_c}{2} \left(\frac{r}{D_{SAR}} \right) \quad (7.16)$$

By combining (7.15) with (7.16) we can solve for a limit on down-range to the scene

$$r \leq \frac{4\Delta CR^2}{\lambda_c} \quad (7.17)$$

As resolution improves (decreases), DBS is limited to short-range collections. For example, at 10 GHz (0.03 m wavelength) and a 15 m resolution, DBS is constrained to operate inside 30 km.

This Doppler smearing effect is captured explicitly by the PSR. When going from (7.10) to (7.11) a u^2 term was omitted. When retained, this term results in a quadratic phase history

$$\Delta\psi_{QT} \equiv \frac{4\pi}{\lambda_c} \left(\frac{u^2}{2r} \right) \quad (7.18)$$

where QT denotes *quadratic term*. In [1] it was shown that the SAR collection baseline limit in (7.15) is replicated by limiting the phase deviation of the quadratic term in (7.18) to less than 90° .

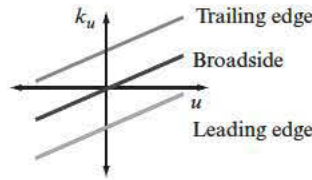


FIGURE 7-9 ■ When the quadratic phase term becomes significant the spatial frequency varies linearly over the SAR dwell.

With the retention of the quadratic contribution the phase progression in (7.11) becomes

$$\Delta\psi = -\frac{4\pi}{\lambda_c}u \left(\frac{x}{r}\right) + \frac{4\pi}{\lambda_c} \left(\frac{u^2}{2r}\right) \quad (7.19)$$

and the corresponding spatial frequency is

$$k_u = -\frac{4\pi}{\lambda_c} \left(\frac{x}{r}\right) + \frac{4\pi}{\lambda_c} \left(\frac{u}{r}\right) \quad (7.20)$$

The spatial frequency (Doppler) is no longer constant but varies linearly over the dwell, as depicted in Figure 7-9. This linear frequency modulation (LFM) is also known as a chirp function.

The quadratic phase term possesses two favorable traits. First, it is not a function of the cross-range location x of the scatterer; all scatterers at the same down-range r are modulated by the same quadratic phase signal. Second, the modulation may be accurately predicted using knowledge of the platform motion (the u history). Therefore, the chirp modulation may be removed by dechirping the raw data by multiplying it by the conjugate of the predicted quadratic phase history. This dechirp function

$$e^{-j\Delta\psi_{QT}(u;r)} \quad (7.21)$$

depends only on the platform location and the distance into the data down-range. This azimuth (AD) operation is depicted in Figure 7-10 and is reminiscent of the analog deramp stage in stretch processing for pulse compression. In older SAR texts [2] it is sometimes referred to as SPECAN (for SPECTral ANalysis). For narrow swaths (limited down-range image extents) a single dechirp function may suffice, though in general a different dechirp function may be applied to each range bin, consistent with the dependence on r in (7.18) [3].

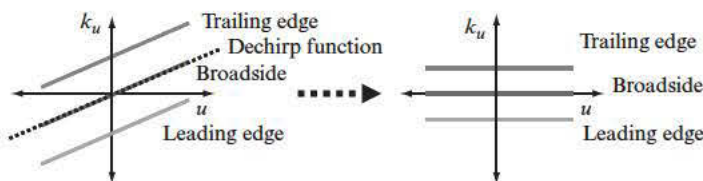
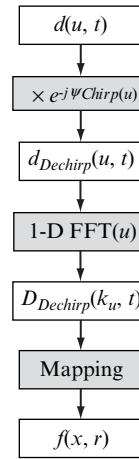


FIGURE 7-10 ■ Azimuth dechirp (AD) consists of multiplying the data over the along-track dimension by a phase function equal to the conjugate of the predicted quadratic phase term. AD serves to remove the LFM, yielding signals having constant spatial frequency over the collection.

FIGURE 7-11 ■ DBS-AD is similar to DBS but preprocesses the data to DBS with a dechirp stage.



The flow diagram for DBS with AD appears in the Figure 7-11 and differs from the DBS diagram in Figure 7-6 only in the addition of the dechirp stage. The compensated data is denoted $d_{Dechirp}(u, t)$.

The following simulation parameters were used to demonstrate the effect of the quadratic phase term and the benefit of AD to DBS:

- Carrier frequency = 9.5 GHz
- RF bandwidth = 15 MHz
- Integration angle = 0.1°
- Resolution = 10 m (both dimensions)
- Range to scene center = 50 km
- Scene size = 1 km by 1 km

Diagrams of the collection geometry and the location of point targets are provided in Figure 7-12.

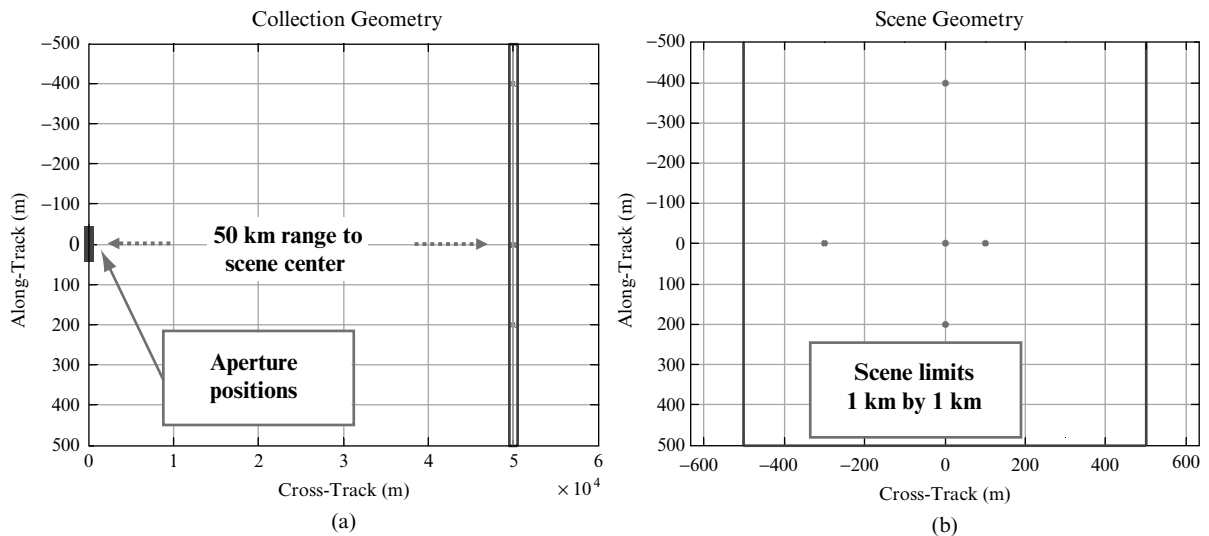


FIGURE 7-12 ■ Simulated DBS-AD geometry emphasizing (a) the overall collection geometry, and (b) point scatterer locations in the scene of interest.

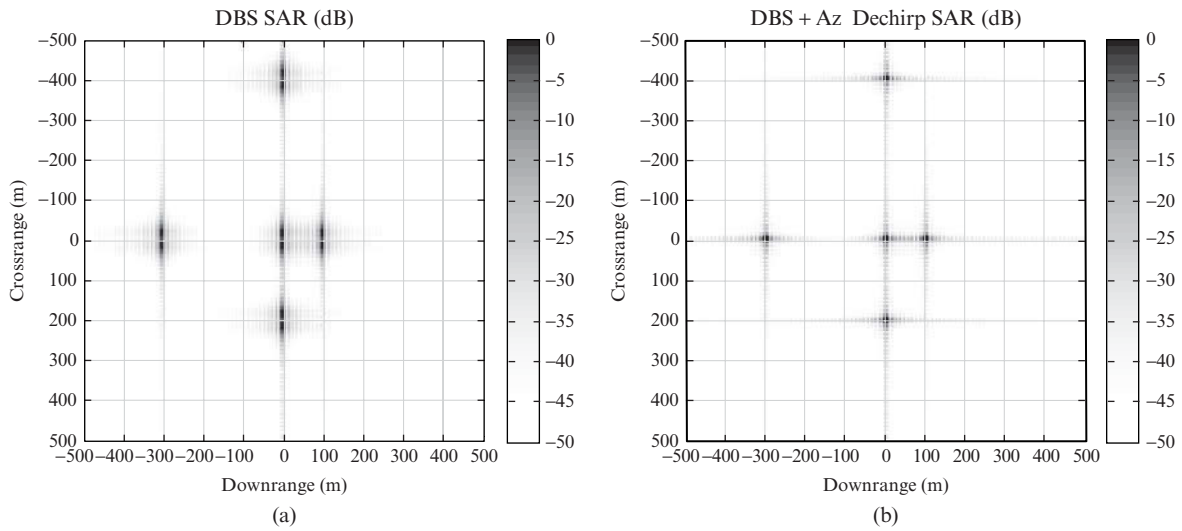


FIGURE 7-13 ■ Simulated imagery for (a) DBS, and (b) DBS-AD.

Figure 7-13a is a DBS image with scatterer returns exhibiting clear smearing in cross-range due to significant quadratic phase over the dwell, while Figure 7-13b demonstrates the benefit of DBS-AD with well-focused point target responses in the image.

The quadratic phase function, the conjugate of which is employed by the AD stage, is shown in Figure 7-14. Image color maps to radians of phase, the maximum phase deviation is about 2π . The ordinate corresponds to platform aperture (along-track sample) number, where the total collection length D_{SAR} was about 80 m. The abscissa is frequency number and covers the 15 MHz waveform bandwidth centered at 9.5 GHz.

A subtle point in Figure 7-14 is that the dechirp function does not depend on frequency. Instead, AD is performed using the quadratic phase as determined by the center RF only. Technically, the chirp function does depend on frequency, as suggested by the wavelength term in (7.18). One can show that applying a frequency-dependent phase correction

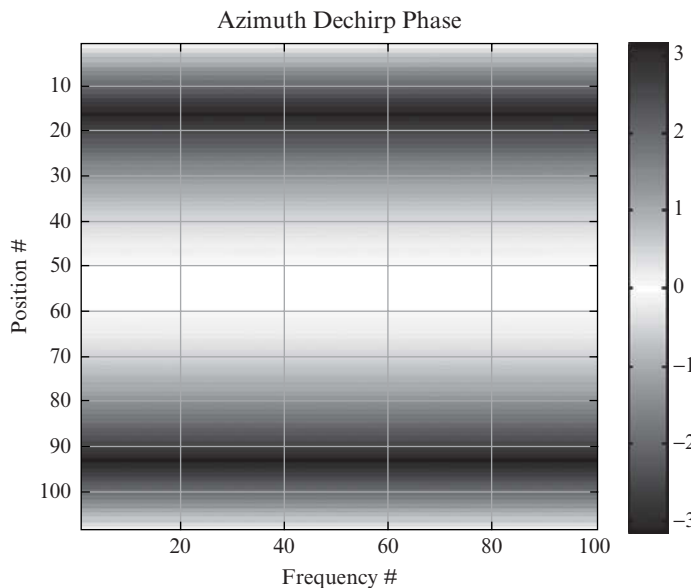


FIGURE 7-14 ■ Quadratic phase term used in the azimuth dechirp stage of DBS-AD.

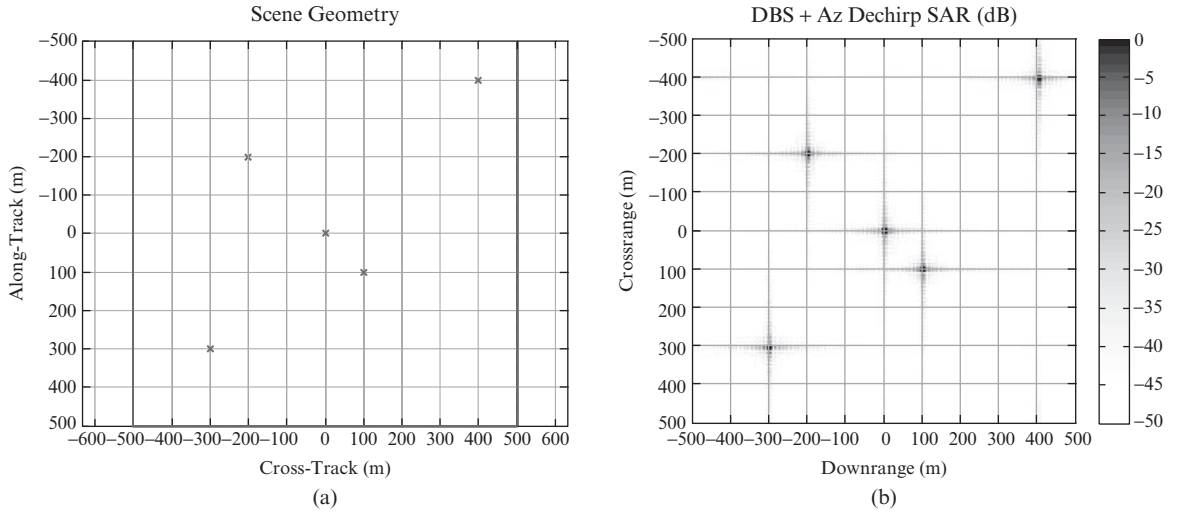


FIGURE 7-15 ■ Modified point scatterer laydown showing (a) scatterer locations, and (b) DBS-AD image.

effectively phase corrects and time aligns the data for a return at the center of the scene. This is tantamount to motion compensation to a point (MCP), a preprocessing act suitable for tomographic (spotlight) image formation techniques. The MCP operation is inconsistent with stripmap imaging, which does not employ a preferred focus point. For stripmap imaging, any preprocessing motion compensation serves only to phase correct and time align the data to a uniformly sampled along-track line, so-called motion compensation to a line (MCL). Therefore, DBS-AD always employs a nondispersive (frequency-independent) phase compensation.

Figure 7-15a exhibits a modified laydown of point scatterers, and Figure 7-15b shows the resulting DBS-AD image. The aim of this laydown is to place each of the five targets into a separate range bin.

Figure 7-16a is an image of the real part of the raw data $d(u, t)$ for the simulated laydown in Figure 7-15. There are five significant range responses, each corresponding to one of the five targets. The undulations between black and white are the bipolar fluctuations between positive and negative voltages by the real part of the along-track phase history and map directly to spatial frequency. These fluctuation rates and, equivalently, the spatial frequencies clearly change over the dwell (along the vertical). Figure 7-14b is the same information after the AD stage. All five scatterers now have a constant bipolar fluctuation rate over the along-track position history of the platform. In other words, the scatterer histories in Figure 7-16b possess constant spatial frequencies over the dwell and so are appropriate for subsequent DBS imaging.

7.3.2 Range Migration Compensation

Adding azimuth dechirp allows DBS to overcome quadratic phase effects and successfully operate over finer resolutions than it could by itself. Continuing our quest for resolution, the next challenge is scatterer migration across range bins during the SAR collection. In Figure 7-17 the platform is moving from left to right, or, equivalently, the platform may be modeled as stationary with scatterers drifting into and out of the mainbeam from right

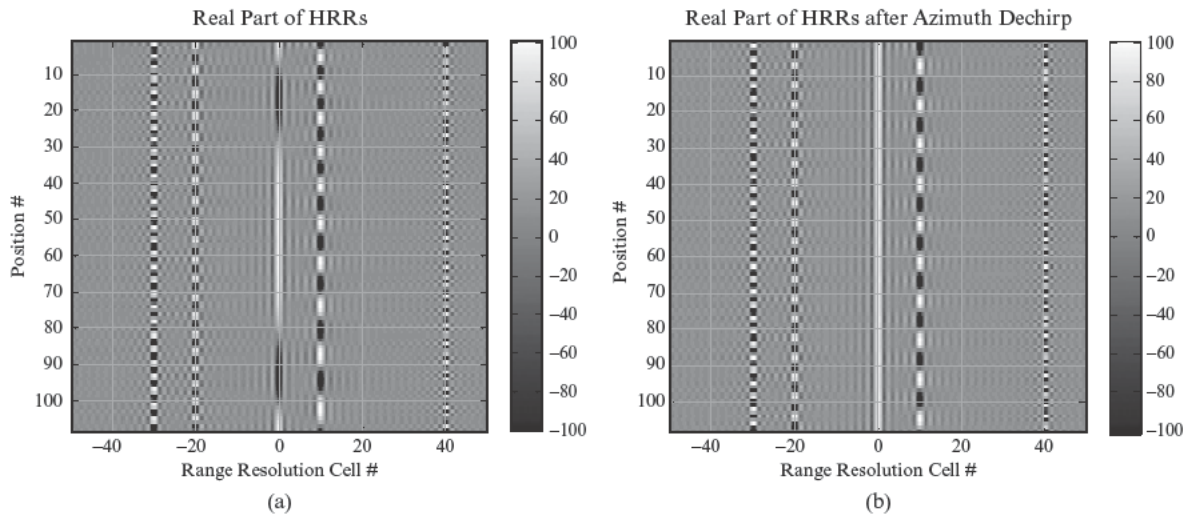


FIGURE 7-16 ■ Real part of the raw data (a) before azimuth dechirp, and (b) after azimuth dechirp.

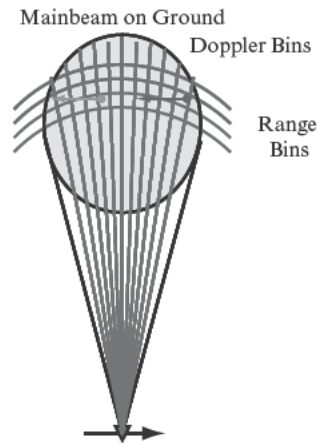


FIGURE 7-17 ■ SAR collection with increasingly fine resolution and consequent scatterer drift over range resolution cells.

to left. Scatterers flow across a significant number of Doppler (cross-range) bins, but AD compensates handsomely. However, with a long dwell and fine range resolution, scatterers begin to migrate over range bins. Recall that while DBS treats these as down-range bins, they are in reality slant-range bins that project onto the ground as circles. Scatterers on the leading edge of the beam (to the right) flow to the left and over range bins toward the radar, whereas scatterers on the trailing edge travel over range away from the radar.

This migration of scatterers over range bins has two deleterious effects on the final DBS image. Obviously, a scatterer will be defocused over the number of bins it migrated through during the collection. More subtly, a scatterer spends just a fraction of the dwell in any given bin, so that for most of the dwell the scatterer is not present in a particular bin. When DBS implements the Fourier transform from u to k_u it will not achieve full resolution on a scatterer and cross-range resolution will be degraded. For example, if a scatterer migrates over three bins its energy is distributed over those three bins. Consequently, down-range resolution is spoiled by a factor of three, and it spends only one-third of the dwell in any given bin so that cross-range resolution is spoiled by a factor of three as well.

FIGURE 7-18 ■
Raw data showing
PSRs for scatterers
on boresight and off
boresight; the solid
rectangular region
denotes data to be
applied to DBS-AD.

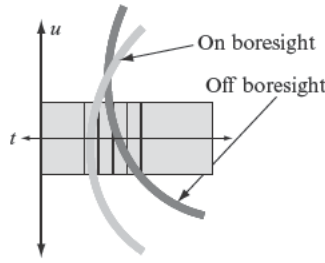


Figure 7-18 depicts this migration effect in the raw data domain $d(u, t)$ for two PSRs: one for a scatterer on the boresight of the antenna (broadside to the platform) and the other from a scatterer off boresight. The solid rectangular region denotes the raw data to be applied to DBS-AD image formation. Range migration is more significant for scatterers away from broadside and is most severe for scatterers at the leading and trailing edge of the beam. If the acquisition coordinate system is aligned so that the center of the dwell $u = 0$ is the same as the center of the scene $x = 0$ (i.e., $u = x$), then range migration is most severe for scatterers having large positive and negative x values.

We may predict the degree of range migration by starting with the PSR in its slant-range form in (7.5) and retaining only the linear portion of the approximation in (7.8)

$$R(u; x, r) \approx -u \frac{x}{r} \quad (7.22)$$

If the platform moves from $u = -D_{SAR}/2$ to $u = +D_{SAR}/2$, the total slant-range change is $x(D_{SAR}/r)$.

A reasonable constraint on range migration is some fraction of the range resolution cell size. A common limit on migration is half of the range resolution, which yields

$$x \frac{D_{SAR}}{r} \leq \frac{\Delta R}{2} \quad (7.23)$$

The requirement in (7.23) places a constraint on the imaged scene half-size as given by x_S

$$x_S \leq \frac{\Delta R}{2} \left(\frac{r}{D_{SAR}} \right) \quad (7.24)$$

Using (7.16) we can rewrite (7.24) in terms of cross-range resolution

$$x_S \leq \frac{\Delta R \Delta CR}{\lambda_c} \quad (7.25)$$

For example, with a 10 m resolution in both dimensions at a 10 GHz RF (0.03 m wavelength), the half-extent of the scene is limited to 3,333 m, for a total extent of 6,666 m.

Figure 7-19 is a simplified representation of Figures 7-17 and 7-18 and highlights the response of scatterers at the leading and trailing edges of the beam. The challenge of range migration compensation is readily apparent: its effects are spatially variant, so different scatterers require different correction functions. This is in stark contrast to quadratic phase effects, where the same dechirp function is good for all scatterers.

On the other hand, Figure 7-19 suggests a possible solution to this problem. A scatterer migrating inbound will have a negative phase progression over the dwell, while an

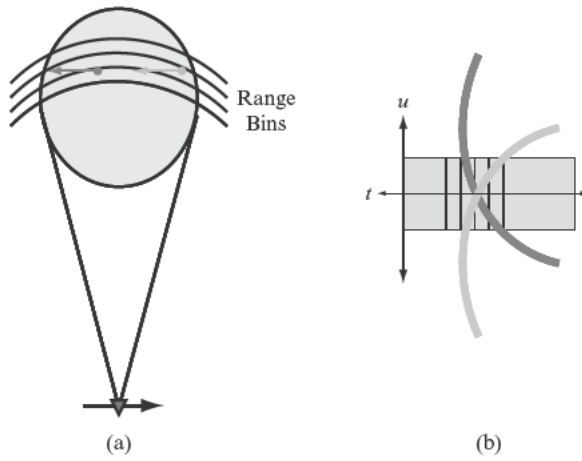


FIGURE 7-19 ■ Range migration for a scatterer at the leading and trailing edge of the beam (a) in an overhead view, and (b) as PSRs in the raw data.

outbound scatterer will have a positive phase progression. In other words, the scatterer on the leading edge of the beam in Figure 7-19 will exhibit high negative spatial frequencies, whereas the scatterer on the trailing edge will exhibit high positive spatial frequencies. Because there appears to be a unique mapping from range migration magnitude and direction to spatial frequency, a compensation may be developed and applied in the k_u domain instead of the u domain where AD operates.

We begin with the PSR and approximate the quadratic as before by assuming $(u - x) \ll r$

$$R = \sqrt{(u - x)^2 + r^2} \approx r + \frac{(u - x)^2}{2r} \quad (7.26)$$

First, isolate the change in slant range with respect to r , the slant range at the closest point of approach, and call it the range migration, R_{Mig} ,

$$R_{\text{Mig}} = R - r \approx \frac{(u - x)^2}{2r} \quad (7.27)$$

Next, use (7.27) to find the phase due to range migration

$$\psi_{\text{Mig}} = 4\pi \frac{R_{\text{Mig}}}{\lambda} = 2\pi \frac{(u - x)^2}{\lambda r} \quad (7.28)$$

Note that, in contrast to azimuth dechirp, we use the general wavelength λ , not just the wavelength at the center frequency, λ_c . It turns out the compensation for range migration includes a dispersive (frequency-dependent) term.

Then, find the spatial frequency due to migration

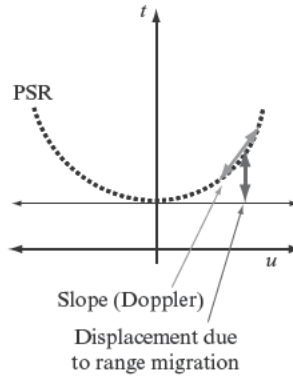
$$k_{u,\text{Mig}} \equiv \frac{\partial \psi_{\text{Mig}}}{\partial u} = \frac{4\pi (u - x)}{\lambda r} \quad (7.29)$$

Finally, we can tie phase offset to spatial frequency by exploiting the $(u - x)$ term (7.28) and (7.29) have in common:

$$\psi_{\text{Mig}} = \frac{\lambda r}{8\pi} k_{u,\text{Mig}}^2 \quad (7.30)$$

FIGURE 7-20 ■

There is a one-to-one mapping from displacement due to range migration and instantaneous spatial frequency (Doppler).



Equation (7.30) proves our hypothesis that range migration (captured here by a phase offset) is tied to the slope of the phase progression (Doppler or spatial frequency). This one-to-one mapping in the raw data is underscored in Figure 7-20.

Replacing wavelength with radian frequency ($\lambda = 2\pi c/\omega$) allows to express (7.30) in the RF domain

$$\psi_{\text{Mig}} = \frac{cr}{4} \left(\frac{1}{\omega} \right) k_{u,\text{Mig}}^2 \quad (7.31)$$

The RMC procedure begins by Fourier transforming the raw data in both dimensions, from u to k_u and from t to ω . The first transform is given by (7.13), and the second is

$$D(k_u, \omega) = \int_{-\infty}^{\infty} D(k_u, t) e^{-j\omega t} dt \quad (7.32)$$

Then the frequency-domain data are corrected by multiplying by the conjugate of the phase term in (7.31)

$$D_{\text{RMC}}(k_u, \omega) = D(k_u, \omega) e^{-j\psi_{\text{Mig}}} \quad (7.33)$$

It happens that the compensation term in (7.31) is too effective, as it strips away the phase information DBS uses to map spatial frequency to cross-range. To correct this deficiency we modify (7.31) by removing the correction at the center frequency, ω_c .

$$\psi_{\text{Mig}} = \frac{cr}{4} \left(\frac{1}{\omega} - \frac{1}{\omega_c} \right) k_{u,\text{Mig}}^2 \quad (7.34)$$

It is seen that the phase compensation in (7.34) is zero if RF equals the center frequency or if the spatial frequency equals zero.

The flow diagram in Figure 7-21 shows the RMC procedure in the left-side column, including the 2-D inverse Fourier transform to take the compensated data out of the frequency domain and back into the space-time (u, t) domain.

$$d_{\text{RMC}}(u, t) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} D_{\text{RMC}}(k_u, \omega) e^{+jk_u u} e^{+j\omega t} dk_u d\omega \quad (7.35)$$

Thereafter, AD and DBS proceed as described in earlier sections.

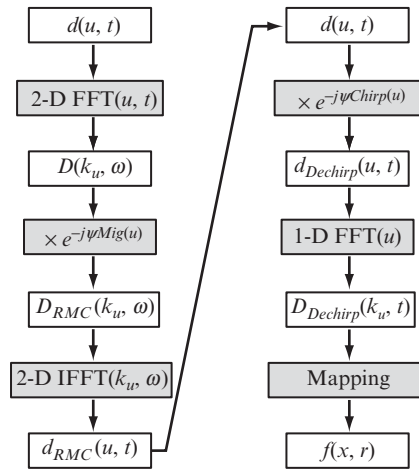


FIGURE 7-21 ■ DBS with RMC is similar to DBS-AD but preprocesses the data with an RMC stage.

The following simulation parameters were used to demonstrate the effect of the range migration and the benefit of RMC to DBS-AD:

- Carrier frequency = 250 MHz
- RF bandwidth = 7.5 MHz
- Integration angle = 1.7°
- Resolution = 20 m (both dimensions)
- Range to scene center = 50 km
- Scene size = 2 km by 2 km

The use of a low carrier frequency is designed to stress the scene size, as (7.25) indicates cross-range extent will be limited when operating with long wavelengths. Diagrams of the collection geometry and the location of point targets are provided in Figure 7-22.

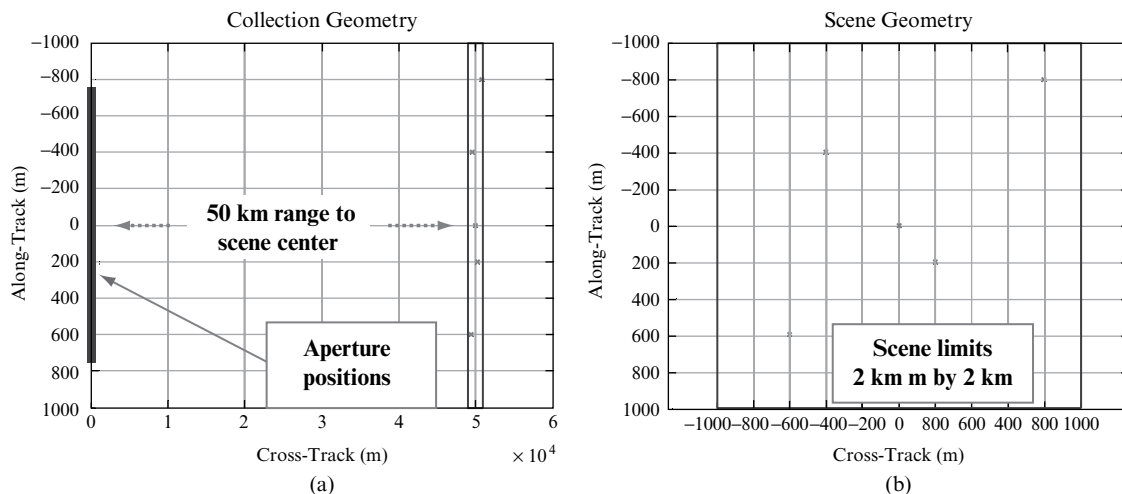


FIGURE 7-22 ■ Simulated DBS-AD-RMC geometry emphasizing (a) the overall collection geometry, and (b) point scatterer locations in the scene of interest.

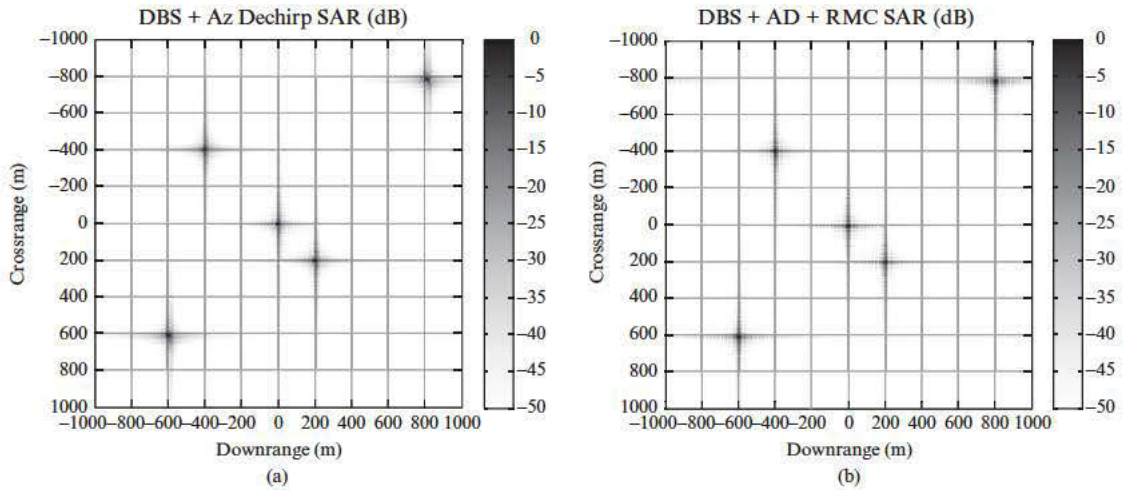


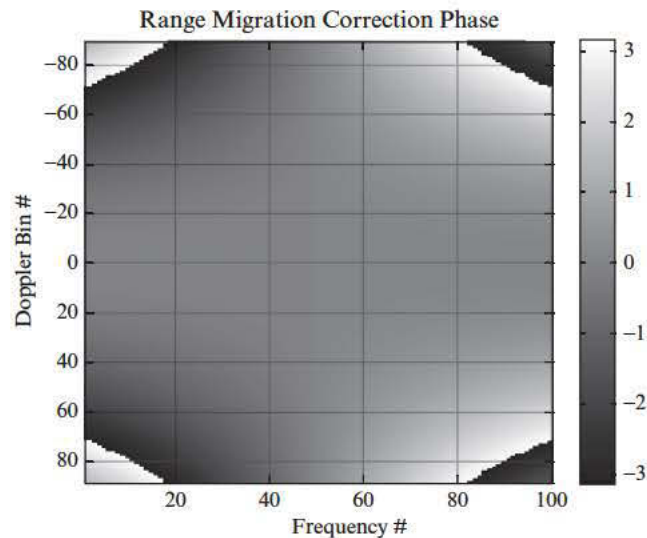
FIGURE 7-23 ■ Simulated imagery for (a) DBS-AD, and (b) DBS-AD-RMC.

Figure 7-23a is a DBS-AD image with a well-focused target at scene center but other scatterers defocused in both down-range and cross-range, consistent with range migration over the dwell. Targets at large positive and negative cross-range locations are more severely degraded, which makes sense since these sources are away from broadside and experience increase range migration. Figure 7-23b demonstrates the benefit of RMC to DBS-AD with all point targets in the image in focus.

The RMC function in (7.34), the conjugate of which is employed by the RMC stage, is shown in Figure 7-24. Image color maps to radians of phase; the maximum phase deviation is just under 2π . The abscissa is frequency number over the 7.5 MHz waveform bandwidth centered at 250 MHz, and the ordinate is spatial-frequency bin numbered with respect to zero frequency.

Figure 7-25a shows range profiles for the five scatterers as a function of platform position with arrows denoting the closest point of approach for each of the hyperbolic

FIGURE 7-24 ■ RMC phase function versus spatial frequency and RF.



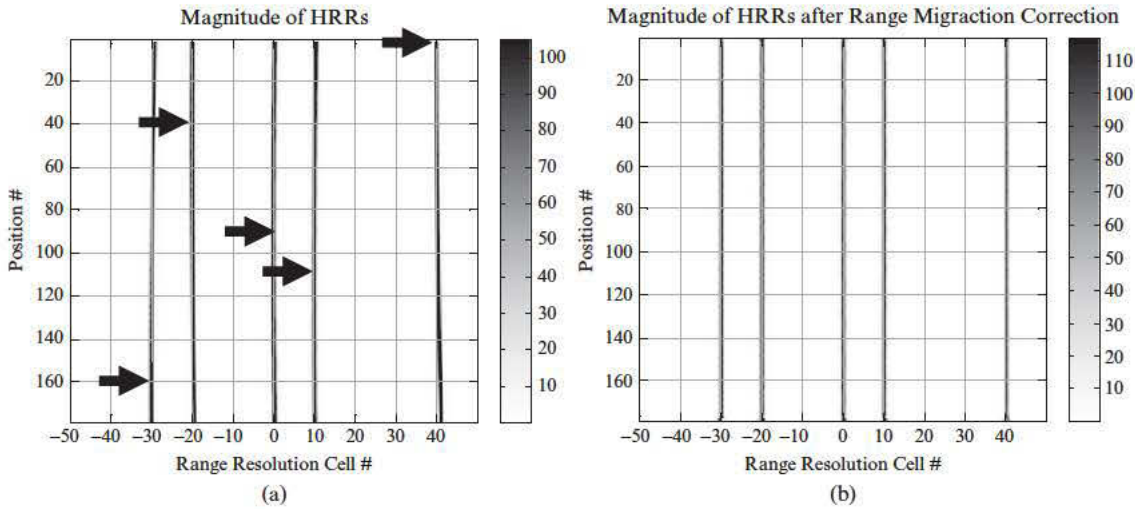


FIGURE 7-25 ■ Simulated range profiles versus along-track position (a) before RMC, and (b) after RMC.

PSRs. Range migration is most significant for targets having the closest point of approach toward the beginning or end of the collection. Compensation for migration in this domain is impossible, as shifting range profiles to align the return for one scatterer serves to exacerbate the migration of other scatterers. Figure 7-25b shows an absence of migration on all five targets after RMC is implemented in the 2-D frequency domain using the function in Figure 7-24.

The response of the scatterers over spatial-frequency/Doppler and range appear in Figure 7-26a. In contrast to the aperture position responses in Figure 7-25, the spatial-frequency responses all exhibit the same closest point of approach, which is highlighted by the arrows and occurs at $k_u = 0$. Repairing migration is feasible in this domain: we can imagine range-shifting the profiles, with increasing delays as spatial frequency becomes more positive and negative, to remove migration over range. The RMC function in Figure 7-24 accomplishes this alignment by applying phase progressions over RF. The migration-free product is given in Figure 7-26b. The concept of fixing migration in the spatial-frequency domain will be revisited in the next section with RDA.

It happens that RMC works well even when all scatterers in the scene have significant hyperbolic histories and consequent range migration. This robustness is demonstrated with an X-band simulation using the following parameters:

- Carrier frequency = 9.5 GHz
- RF bandwidth = 150 MHz
- Integration angle = 0.9°
- Resolution = 1.0 m (both dimensions)
- Range to scene center = 25 km
- Scene size = 100 m by 100 m

Diagrams of the collection geometry and the location of point targets appear in Figure 7-27.

FIGURE 7-26 ■
Simulated range
profiles versus
spatial frequency
(a) before RMC, and
(b) after RMC.

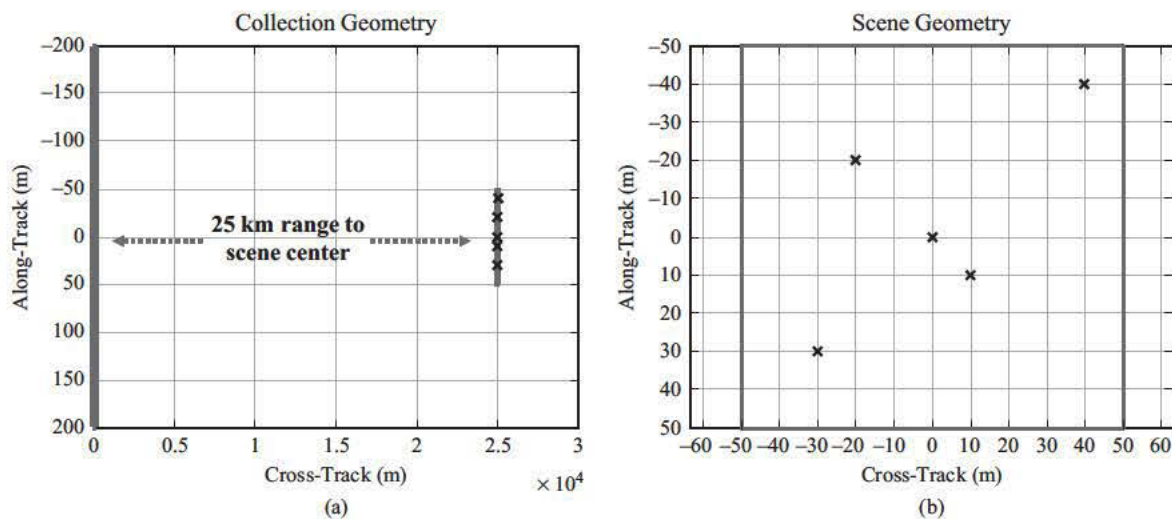
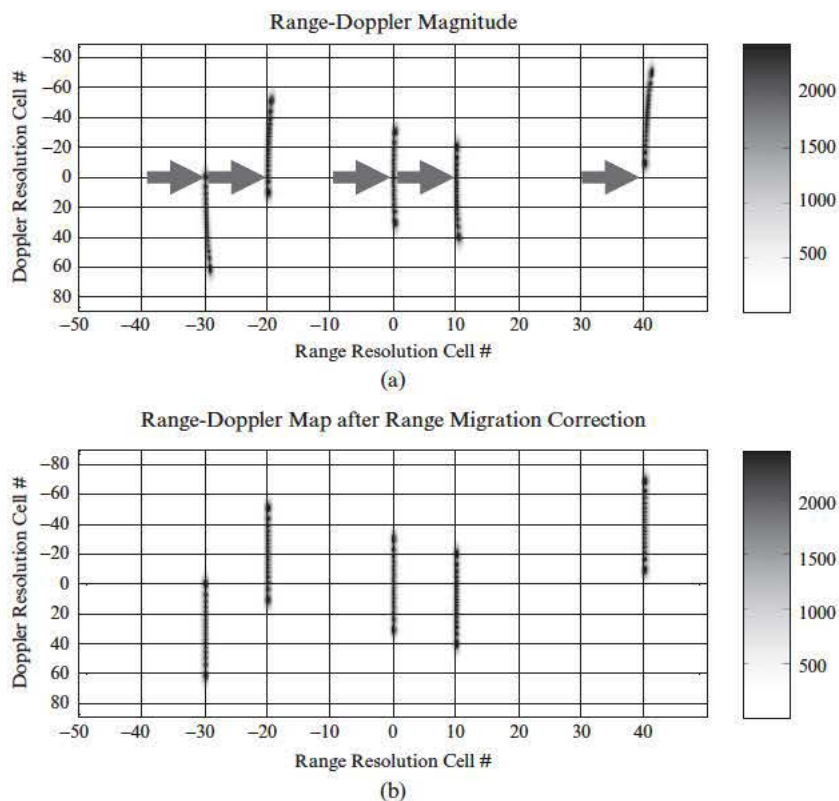


FIGURE 7-27 ■ Simulated DBS-AD-RMC geometry emphasizing (a) the overall collection geometry, and (b) point scatterer locations in the scene of interest.

The DBS-AD image is provided in Figure 7-28a and shows that all the scatterers are defocused in both down-range and cross-range, suggesting range migration is a problem even for scatterers at broadside. Figure 7-28b contains the DBS-AD-RMC image with all scatterers well focused.

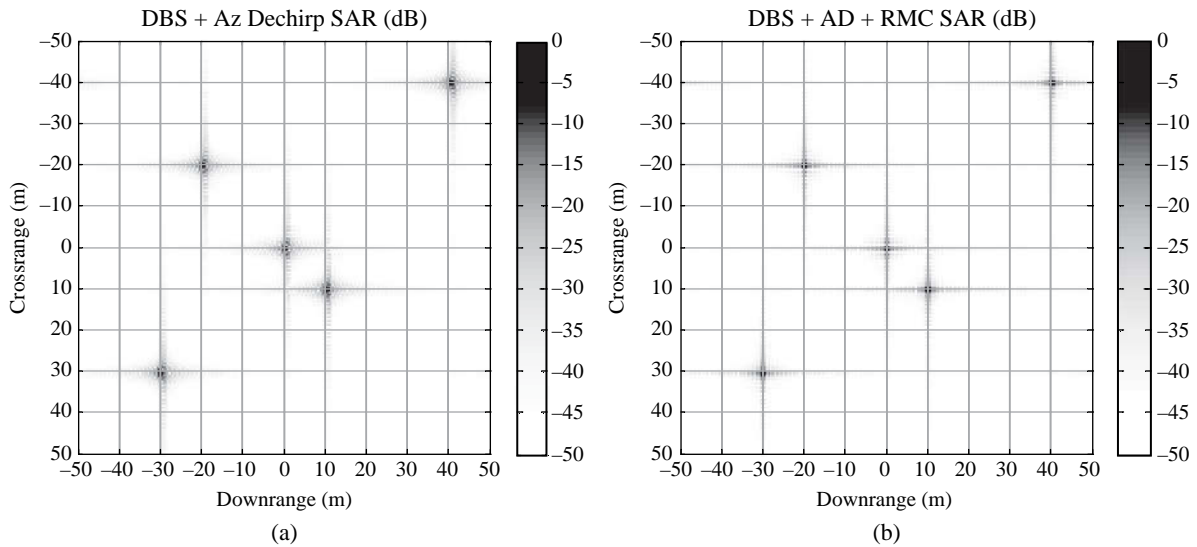


FIGURE 7-28 ■ Simulated imagery for (a) DBS-AD, and (b) DBS-AD-RMC.

7.3.3 Notes on Doppler Beam Sharpening

As capable as the DBS family of techniques may appear, they suffer from a number of drawbacks and are rarely employed in high-fidelity imaging. For example, the output image is warped because the algorithm operates on circles of slant range and hyperbolic iso-Doppler lines. The resulting geometric distortion makes geographic registration of DBS imagery difficult. In addition, imaging at very fine resolutions over wide scenes generates higher-order phase and migration histories the DBS paradigm cannot readily accommodate.

Worse yet, DBS has fatal flaws for stripmap applications. Stripmap acquisitions are typically long affairs with many scatterers moving through the mainbeam of the real antenna over an extended tract of terrain. Imagine a fairly benign collection wherein higher-order phase effects and range migration are negligible so that DBS-AD will suffice for image formation. We might begin by processing a subset of the data, applying a block of along-track records to DBS-AD. If the block size is equal to the illumination interval for a point target, DBS-AD will focus a return at broadside to full resolution. However, returns at other cross-range locations will have been illuminated over only a fraction of the block, so their returns will be focused to a degraded cross-range resolution. Block size could be increased, thereby ensuring full resolution over a finite along-track extent, but the problem comes in the azimuth dechirp stage. Referring to Figure 7-10, dechirp maps along-track position to spatial frequency. Larger block sizes include more along-track returns and generate higher and lower frequencies after dechirp. However, the data are sampled discretely along-track, so these signals will soon begin to wrap in frequency. The end product will be highly aliased in cross-range.

Finally, the DBS approach is rather ad hoc: as damaging affects appear we tack on additional stages to fix them. Such a strategy is not sustainable. The next section develops a family of techniques that are much more intellectually pleasing.

7.4 | RANGE-DOPPLER ALGORITHMS

In developing the image formation algorithms in this section we will directly exploit the PSR as given in (7.7). Recall the properties of the PSR while referring to Figure 7-29:

The PSR is, in general, a two-dimensional function (spatial and temporal, u and t).

- The PSR is not separable in u and t .
- The PSR has a hyperbolic form for linear collections.
- The PSR form varies with scatterer cross-track (down-range) location.
- The PSR form is invariant with scatterer along-track (cross-range) location.

Our goal is to build a matched filter for imaging using the PSR. To that end, we begin with a review of the matched filter.

7.4.1 Matched Filtering

Many radar signal processing applications involve a known one-dimensional impulse response $s(t)$. For example, in pulse compression applications the known impulse response is the transmitted waveform. Recorded data, $d(t)$, consist of the superposition of time-delayed, amplitude-scaled, and phase-offset replicas of the impulse response. Figure 7-30 exhibits a stylized depiction of an impulse response and a data record.

Real data always contain some thermal noise and often other undesirable components. The data may be filtered to suppress interference and accentuate the known signals. Filtering may be described using the temporal impulse response of the filter, $h(t)$, and convolving the filter with the data to generate an output, $y(t)$

$$y(t) = d(t) * h(t) \equiv \int_{-\infty}^{\infty} d(\tau) h(t - \tau) d\tau \quad (7.36)$$

FIGURE 7-29 ■ The PSR varies with down-range but not cross-range.

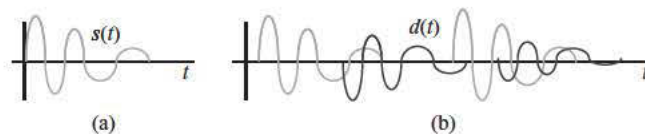
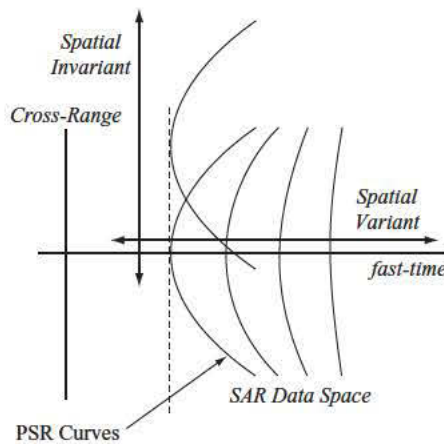


FIGURE 7-30 ■ One-dimensional signals for (a) an impulse response, and (b) a data record.

where $*$ denotes convolution. Note one feature of convolution in (7.36): a change in sign on the dummy variable of integration between the two terms in the integral.

Alternatively, filtering may be implemented in the frequency domain. Using the Fourier transform to find the frequency response of the data, $D(\omega)$,

$$D(\omega) = \int_{-\infty}^{\infty} d(t)e^{-j\omega t} dt \quad (7.37)$$

and filter, $H(\omega)$,

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t} dt \quad (7.38)$$

filtering is simply a multiplicative process with output, $Y(\omega)$.

$$Y(\omega) = D(\omega)H(\omega) \quad (7.39)$$

The matched filter is designed to maximize the signal-to-noise ratio (SNR) on a known signal. The matched filter, $h_0(t)$, for a signal, $s(t)$, in additive white Gaussian noise (AWGN) is

$$h_0(t) = s^*(-t) \quad (7.40)$$

The matched filter is simply the conjugate of the time-reversed signal to which the filter is matched. Applying the matched filter in (7.40) to the convolution in (7.36) yields

$$y(t) = \int_{-\infty}^{\infty} d(\tau) s^*(\tau - t) d\tau \quad (7.41)$$

In contrast to the convolution equation, the dummy variable of integration has the same sign in both terms, and one term is conjugated with respect to the other. These are the hallmarks of not convolution but correlation. The matched filter output, then, is the correlation of the data with the conjugate of the signal of interest.

The Fourier transform of the matched filter, $H_0(\omega)$, can be found, through a reordering of the conjugation operation on the Fourier integral, to be

$$H_0(\omega) = S^*(\omega) \quad (7.42)$$

Applying (7.42) to (7.39) yields the frequency-domain output of the matched filter

$$Y(\omega) = D(\omega)S^*(\omega) \quad (7.43)$$

In anticipation of a SAR application we extend the matched filter development to two dimensions. The signal, $s(u, t)$, constitutes the known system impulse response; in SAR, $s(u, t)$ corresponds to the PSR. The matched filter in AWGN is the conjugate of the signal reversed in both dimensions

$$h_0(u, t) = s^*(-u, -t) \quad (7.44)$$

and in the frequency domain it is the conjugate of the signal

$$H_0(k_u, \omega) = S^*(k_u, \omega) \quad (7.45)$$

Time-domain matched filtering is a 2-D convolution

$$y(u, t) = d(u, t) *_{u,t} s^*(-u, -t) \quad (7.46)$$

which reduces to a 2-D correlation with the known signal

$$y(u, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(v, \tau) s^*(v - u, \tau - t) dv d\tau \quad (7.47)$$

Note the use of two dummy variables of integration, v and τ . The frequency-domain output is simply the product of the data and the conjugate of the signal

$$Y(k_u, \omega) = D(k_u, \omega) S^*(k_u, \omega) \quad (7.48)$$

The mathematics become more complicated when the signal is spatially variant; that is, the impulse response is a function of the target location. In that case we define the known signal impulse response as $s(u, t; u', t')$, where u and t are the support for the signal in the raw data as before, and u' and t' denote the signal source (e.g., scatterer location). The convolution integral takes the form

$$y(u, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(v, \tau) s^*(v, \tau; u, t) dv d\tau \quad (7.49)$$

Figure 7-31 highlights the role of the variables in (7.49); *output location* can be thought of as an image pixel. Instead of generating just one matched filter and sliding it through the data in a correlation process, a different matched filter must be generated for every output (u, t) . This may be computationally intensive, depending upon how difficult it is to generate the impulse response. Worse yet, there is no frequency-domain expression for spatially variant filtering.

7.4.2 Range Stacking Algorithm

We can build an image formation algorithm by rewriting the spatially variant form of the matched filter in (7.49) using source and output locations given by down-range and cross-range (x, r) and data location given by (u, t)

$$f(x, r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(u, t) h^*(u, t; x, r) du dt \quad (7.50)$$

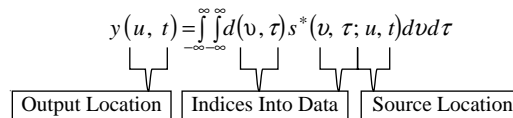


FIGURE 7-31 ■ Two-dimensional matched filtering for a spatially variant signal.

and substituting in the PSR in (7.7) as the signal of interest for the matched filter

$$f(x, r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(u, t) \delta_D \left(t - \frac{2}{c} \sqrt{(u-x)^2 + r^2} \right) dudt \quad (7.51)$$

A close examination of the dependence on u and x of the two terms inside the integrals and the output term suggests that (7.51) has the form of a correlation in u . Rewriting it as an explicit convolution in u with the matched filter yields

$$f(x = u, r) = \int_{-\infty}^{\infty} d(u, t) *_u h(-u, t; r) dt \quad (7.52)$$

where the cross-range x in the output image $f(x, r)$ is found by setting it equal to the along-track u output of the right side of (7.52). We are taking advantage of the fact that the PSR, though spatially variant in cross-track (down-range) r , is spatially invariant along-track (cross-range), over u and x . For one down-range location, r_0 , (7.52) becomes

$$f(x = u, r_0) = \int_{-\infty}^{\infty} d(u, t) *_u h(-u, t; r_0) dt \quad (7.53)$$

The same PSR is good for all cross-range pixels at this down-range location.

The matched filter image former has an outer loop over down-range r . At each down-range r_0 of interest, we begin by forming the reference PSR for that down-range $h(u, t; r_0)$ from (7.7) and setting $x = 0$

$$h(u, t; r_0) = \delta \left(t - \frac{2}{c} \sqrt{u^2 + r_0^2} \right) \quad (7.54)$$

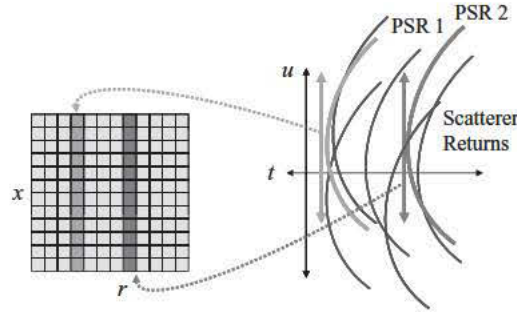
Next we perform the one-dimensional convolution over $u - x$ using (7.53)

$$\begin{aligned} f(x, r_0) &= \int_{-\infty}^{\infty} d(u, t) *_u h^*(-u, t; r_0) dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(u, t) \delta \left(t - \frac{2}{c} \sqrt{(u-x)^2 + r_0^2} \right) dudt \end{aligned} \quad (7.55)$$

Note that, although the convolution is one-dimensional, the data and PSR are still two-dimensional.

Figure 7-32 depicts this matched filtering in stylized form. The output image pixels over x and r are at left, and the raw data over u and t are to the right. A new PSR is generated for each down-range set of cross-range pixels (a column in the image). The PSR is convolved with the data by moving it vertical through the data; at each vertical offset the PSR and data are cross-multiplied and summed in both dimensions, and the scalar output placed into the cross-range pixel corresponding to the vertical offset. If the PSR aligns with the return from a strong target the PSR will coherently reinforce that signal history during the two-dimensional summation, producing a strong output at that pixel location. Once all offsets are evaluated, the convolution is complete and a column of image pixels are formed. Then the process repeats for another down-range location having a different PSR.

FIGURE 7-32 ■
Depiction of
matched filtering of
data for RSA image
formation.



As discussed in the previous section, convolution filtering can also be realized as a multiplication in the frequency domain. Specifically

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(\omega) e^{j\omega t} d\omega \quad (7.56)$$

where $Y(\omega)$ is given by (7.43) and (7.37), and $S(\omega)$ is simply

$$S(\omega) = \int_{-\infty}^{\infty} s(t) e^{-j\omega t} dt \quad (7.57)$$

While multiplication is much easier than convolution, many operations are involved in the three Fourier transforms: two to take both signals—the data $y(t)$ and the matched filter $s*(-t)$ —into the frequency domain, and a final Fourier transform to take the product from the frequency domain back to the time domain. However, these transforms can be accomplished very efficiently through the fast Fourier transform (FFT). For a signal of length N the number of complex operations to support a convolution is on the order of N^2 , whereas for frequency-domain multiplication achieved with the FFT the number of operations scales as $3N \log_2 N$ (one factor of $N \log_2 N$ for each of the three FFTs). For example, for $N = 2,048$, approximately four million complex adds and multiplies are required to convolve two N -length sequences. In contrast, only 67,500 complex adds and multiplies are needed for frequency-domain-based convolution, a decrease in computations of a factor of 60.

Caveats abound regarding the previously taken position. For example, we typically desire linear convolution, but frequency-domain multiplication realizes circular convolution, so care must be taken in zero padding and indexing into the data. In addition, modern computing architectures are driven not just by the number of operations but by input-output and memory considerations, parallelization potential, and a host of related implementation issues. Nevertheless, FFT-based filtering tends to be far more efficient than brute-force convolution, especially for SAR applications. Indeed, a common theme across SAR image formation techniques is the quest to format the data in such a way that an FFT will suffice to form the final image.

Therefore, we reformulate the convolution in u in (7.55) as a multiplication in k_u . First, the data are transformed to the spatial-frequency domain using (7.13). We then loop over down-range r ; for each r_0 the reference PSR is found from (7.54) and is transformed

to the spatial-frequency domain with

$$H(k_u, t; r_o) \equiv \int_{-\infty}^{\infty} h(u, t; r_o) e^{-jk_u u} du \quad (7.58)$$

Then the convolution over u is realized as the inverse Fourier transform over k_u of the product of the data and PSR

$$\begin{aligned} f(x = u, r_o) &= \int_{-\infty}^{\infty} (d(u, t) *_u h^*(-u, t; r_o)) dt \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} D(k_u, t) H^*(k_u, t; r_o) e^{jk_u u} dk_u \right) dt \end{aligned} \quad (7.59)$$

The order of integration in (7.59) may be reversed without error

$$f(x = u, r_o) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} D(k_u, t) H^*(k_u, t; r_o) dt \right) e^{jk_u u} dk_u \quad (7.60)$$

By summing over time first the two-dimensional product is reduced to a one-dimensional signal, so the final inverse Fourier transform is a much more reasonable one-dimensional operation.

This image formation method is known as the range stacking algorithm (RSA) [4] because it outputs one down-range set of cross-range bins at a time and stacks them next to one another, as Figure 7-33 suggests. The RSA procedure is summarized in flow diagram form in Figure 7-33.

RSA is an exact image former in that it makes no approximations to the PSR, in contrast to the DBS family of techniques. It is more computationally efficient than backprojection, the brute-force pixel-by-pixel approach suggested by (7.51) and described in some detail in [1], because convolution is implemented in the frequency domain for one of the two dimensions. Finally, it fully accounts for the down-range variation in the PSR by explicitly building a different matched filter for each down-range set of pixels.

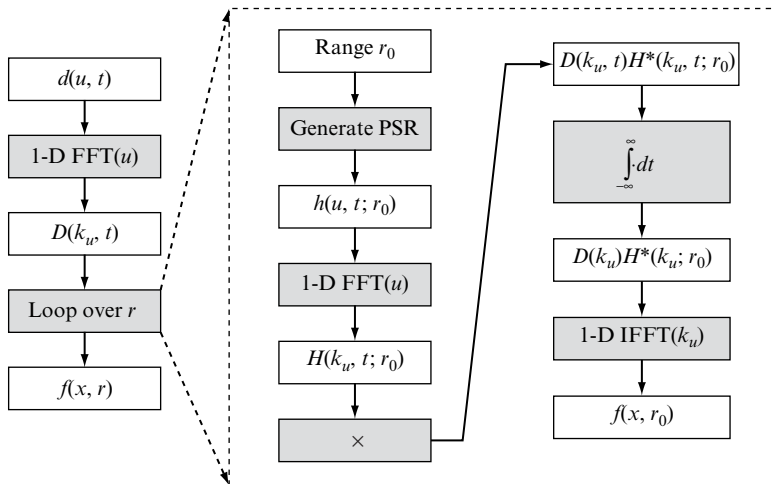


FIGURE 7-33 ■
Signal flow diagram
for the range
stacking algorithm.

Despite these advantages, RSA has not seen widespread application to stripmap SAR imaging. To some extent, RSA is the worst of two algorithms. For example, to implement the FFT-based convolution in RSA, aperture samples must fall on a line and be uniformly spaced. Albeit inefficient, backprojection readily accommodates more general collection geometries. And should the slow-time samples be linear and uniformly spaced, RMA achieves the same level of accuracy as RSA but benefits from a two-dimensional FFT efficiency.

We use RSA as a stepping-stone to the RDA family of techniques, which have seen widespread application to stripmap SAR imaging. And RDA, in turn, is used as a stepping stone to RMA.

7.4.3 Range-Doppler Algorithm

The philosophy behind RDA is to start with RSA but reuse the same PSR for more than one set of down-range bins.

7.4.3.1 Full Range-Doppler Algorithm

We begin with the PSR in (7.7) and focus on the slant-range term given in (7.5). Let us rewrite the slant range with respect to a down-range r_0

$$R = \sqrt{(u - x)^2 + (r_0 + \Delta r)^2} \quad (7.61)$$

The r_0 might be the center of the down-range swath used for imaging, with Δr providing a differential measure of down-range with respect to this swath center reference.

By expanding the second term

$$R = \sqrt{(u - x)^2 + r_0^2 + 2r_0\Delta r + \Delta r^2} \quad (7.62)$$

and assuming $r_0 \gg \Delta r$, that is, the depth of the imaged swath is much less than the down-range to the swath center, then

$$R \cong \sqrt{(u - x)^2 + r_0^2 + 2r_0\Delta r} \quad (7.63)$$

Now isolate the PSR of the scatterers at the reference range, without further approximation

$$R \cong \sqrt{(u - x)^2 + r_0^2} \sqrt{1 + \frac{2r_0\Delta r}{(u - x)^2 + r_0^2}} \quad (7.64)$$

By again assuming $r_0 \gg \Delta r$ we can eliminate the second radical with the binomial expansion we employed often for DBS, specifically, $(1 + z)^p \rightarrow 1 + pz$ for $z \ll 1$

$$R \cong \sqrt{(u - x)^2 + r_0^2} \left(1 + \frac{r_0\Delta r}{(u - x)^2 + r_0^2} \right) \quad (7.65)$$

and expanding

$$R \cong \sqrt{(u - x)^2 + r_0^2} + \frac{r_0\Delta r}{\sqrt{(u - x)^2 + r_0^2}} \quad (7.66)$$

For the second term if we assume $r_0 \gg |u - x|$, then

$$\begin{aligned} R &\cong \sqrt{(u - x)^2 + r_0^2} + \frac{r_0 \Delta r}{\sqrt{r_0^2}} \\ &\cong \sqrt{(u - x)^2 + r_0^2} + \Delta r \end{aligned} \quad (7.67)$$

Using this expression for slant range the PSR in (7.7) becomes

$$\begin{aligned} h(u, t; x, r_0 + \Delta r) &\cong \delta_D \left(t - \frac{2}{c} \left(\sqrt{(u - x)^2 + r_0^2} + \Delta r \right) \right) \\ &\cong \delta_D \left(\left(t - \frac{2}{c} \Delta r \right) - \frac{2}{c} \sqrt{(u - x)^2 + r_0^2} \right) \end{aligned} \quad (7.68)$$

It took a few approximations, but the PSR in (7.68) is spatially invariant in both dimensions. The PSR for scatterers at different cross-ranges x and down-ranges Δr are simply shifted versions of the PSR for down-range r_0 . Substituting (7.68) for the PSR in the general expression for matched filter imaging in (7.50) yields

$$f(x, r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(u, t) \delta_D \left(\left(t - \frac{2}{c} \Delta r \right) - \frac{2}{c} \sqrt{(u - x)^2 + r_0^2} \right) du dt \quad (7.69)$$

and reindexing the output image with respect to the reference range

$$f(x, \Delta r; r_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(u, t) \delta_D \left(\left(t - \frac{2}{c} \Delta r \right) - \frac{2}{c} \sqrt{(u - x)^2 + r_0^2} \right) du dt \quad (7.70)$$

The equation in (7.70) has the form of a correlation in both u and t , that is,

$$f(x = u, \Delta r = ct/2 - r_0; r_0) = d(u, t) *_{u,t} h(-u, -t; r_0) \quad (7.71)$$

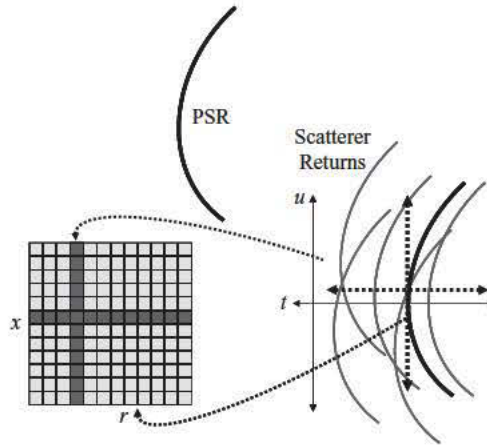
Image formation proceeds by choosing a down-range reference, notionally the center of the swath to be imaged, and generating the PSR for that range using (7.68). Then a two-dimensional convolution is performed with the data and PSR to yield the output image.

Figure 7-34 depicts this matched filtering in stylized form. The output image pixels are at left, and the raw data are to the right. In contrast to RSA shown in Figure 7-32, only one PSR is generated in RDA. This PSR is convolved through the data in both dimensions; at each offset the PSR and data are cross-multiplied and summed in both dimensions, and the scalar output placed into the down-range–cross-range pixel corresponding to the offset. If the PSR aligns with the return from a strong target the PSR will coherently reinforce that signal history during the two-dimensional summation, producing a strong output at that pixel location. Once all convolutional lags are evaluated, image formation is complete.

As discussed in an earlier section, two-dimensional convolution can be realized as a multiplication in the frequency domain. Specifically,

$$y(u, t) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Y(k_u, \omega) e^{+jk_u u} e^{+j\omega t} dk_u d\omega \quad (7.72)$$

FIGURE 7-34 ■
Depiction of
matched filtering of
data for RDA image
formation.



where $Y(k_u, \omega)$ is given by (7.48); together (7.32) and (7.13) yield $D(k_u, \omega)$; more succinctly,

$$D(k_u, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(u, t) e^{-jk_u u} e^{-j\omega t} du dt \quad (7.73)$$

and

$$S(k_u, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(u, t) e^{-jk_u u} e^{-j\omega t} du dt \quad (7.74)$$

Following the development of RSA, multiplication requires far fewer computations than does two-dimensional convolution if the efficient FFT is used to move to and from the frequency domain. Specifically, for a two-dimensional data set and filter of length N on a side, the number of complex operations for convolution scales as N^4 while general frequency-domain convolution requires on the order of $6N^2 \log_2 N$ operations. Because of the sparsity of support for the PSR in the (u, t) domain, the number of operations for convolution typically follows a trend closer to N^3 . Regardless, the advantage of FFT-based convolution is overwhelming. For $N = 2,408$ and N^3 the advantage is a factor of 30, close to the 60 calculated for the analogous RSA example, while for N^4 the advantage explodes to a factor of 63,500.

Motivated by computational savings, we reformulate the convolution in u and t in (7.70) as a multiplication in k_u and ω . First, the data are transformed to the frequency domain using (7.73). Next, a down-range reference r_0 is chosen, the corresponding reference PSR is found from (7.54), and the PSR is transformed to the two-dimensional frequency domain with

$$H(k_u, \omega; r_0) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(u, t; r_0) e^{-jk_u u} e^{-j\omega t} du dt \quad (7.75)$$

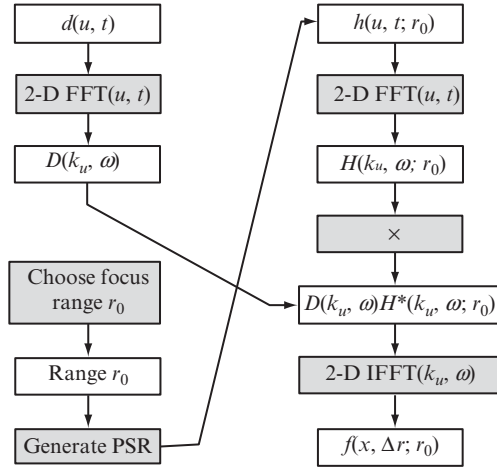


FIGURE 7-35 ■
Signal flow diagram
for the RDA.

Finally, the two-dimensional convolution is effected as a multiplication in (k_u, ω) followed by an inverse Fourier transform

$$f(x = u, \Delta r = ct_2 - r_0; r_0) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (D(k_u, \omega) H^*(k_u, \omega)) e^{jk_u u} dk_u \right) e^{j\omega t} d\omega \quad (7.76)$$

The RDA procedure is summarized in flow diagram form in Figure 7-35.

The following simulation parameters were used to demonstrate RDA:

- Carrier frequency = 9.5 GHz
- RF bandwidth = 300 MHz
- Integration angle = 1.8°
- Resolution = 0.5 m (both dimensions)
- Range to scene center = 7.5 km
- Scene size = 50 m by 50 m

Diagrams of the collection geometry and the location of point targets are provided in Figure 7-36.

Figure 7-37a is a DBS-AD-RMC image with a full set of focused targets, and Figure 7-37b is the analogous image formed with RDA. Both techniques performed well on this data set. However, in comparing Figures 7-21 and 7-35 it is apparent that RDA is a less complicated procedure as it has fewer processing stages.

Figure 7-38 shows the PSR employed as the matched filter in RDA for the simulated data. Figure 7-38a is the phase angle of the frequency-domain rendition of the PSR actually used in (7.76). (The abscissa k_r is a fast-time/RF spatial frequency that corresponds to ω ; specifically, $k_r = 2\omega/c = 2(2\pi/\lambda)$, where the factor of 2 appears because of the radar two-way propagation geometry.) Figure 7-38b is the magnitude of the 2-D Fourier transform of this function; the curved portion of the hyperbolic response is apparent. The discretized function in Figure 7-38a was oversampled in the k_u dimension, resulting in a small degree of zero padding of the hyperbola in cross-range in Figure 7-37b.

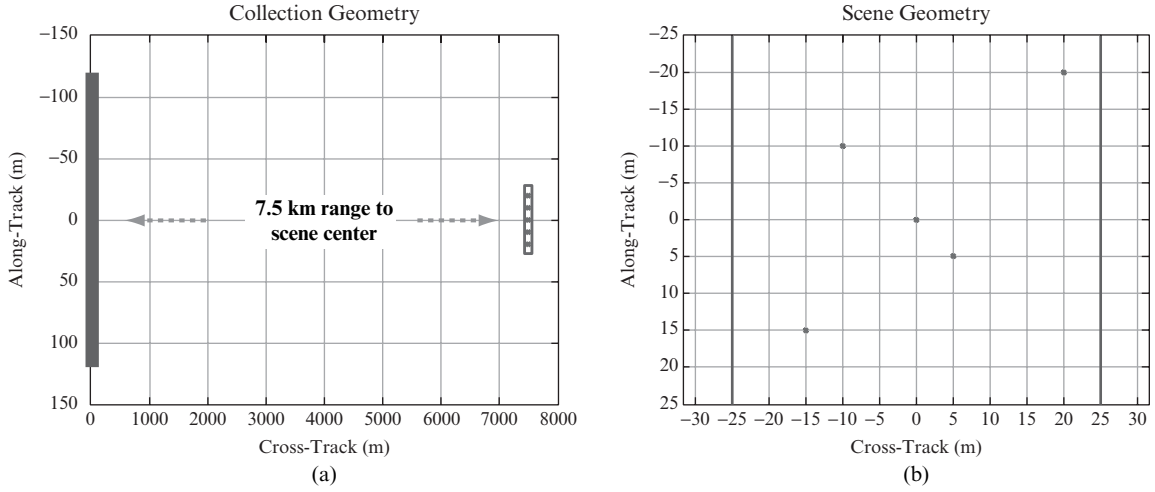


FIGURE 7-36 ■ Simulated RDA geometry emphasizing (a) the overall collection geometry, and (b) point scatterer locations in the scene of interest.

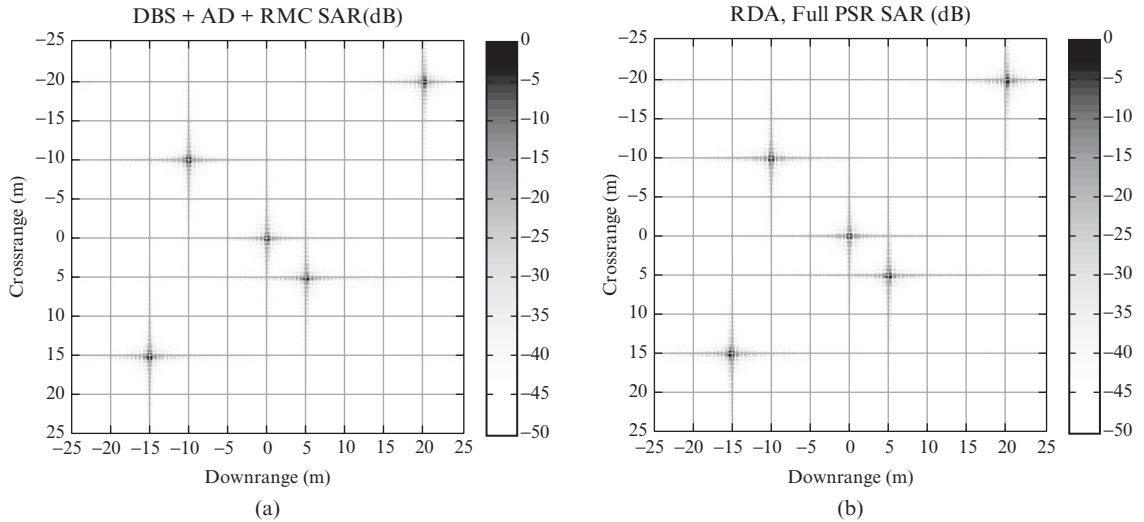


FIGURE 7-37 ■ Simulated imagery for (a) DBS-AD-RMC, and (b) RDA.

7.4.3.2 Depth of Focus

In getting to the spatially invariant PSR employed by RDA a number of approximations were made to the full hyperbolic PSR. These approximations may be fine for a shallow stripmap swath, that is, a limited degree of range offset Δr with respect to the swath center at r_0 . Too deep a swath and the PSR at swath center will be inappropriate for the near and far edges of the image and these regions will not focus properly. The down-range extent over which one PSR is adequate is known as the depth of focus (DOF).

As is often the case in SAR performance analyses, the measure of *adequate* is defined with respect to an allowable phase error. Like the quadratic phase bound suggested in Figure 7-8 and realized in (7.15), let us limit the range error to $\lambda/8$, a two-way phase error of 90° . The range error is given by the difference between the ideal PSR in (7.61) and the

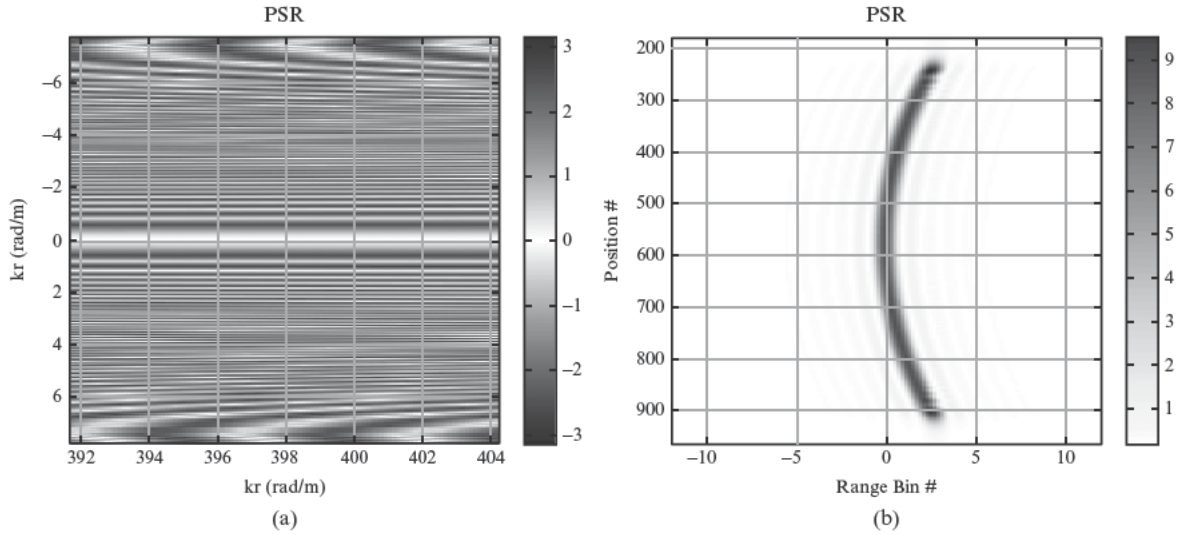


FIGURE 7-38 ■ PSR as used by RDA, (a) two-dimensional spatial-frequency domain, and (b) space-time domain.

approximation in (7.67)

$$\left| \sqrt{(u-x)^2 + (r_0 + \Delta r)^2} - \sqrt{(u-x)^2 + r_0^2} - \Delta r \right| \leq \frac{\lambda}{8} \quad (7.77)$$

Allow $x = 0$; that is, examine two reference PSRs that have been aligned along-track

$$\left| \sqrt{u^2 + (r_0 + \Delta r)^2} - \sqrt{u^2 + r_0^2} - \Delta r \right| \leq \frac{\lambda}{8} \quad (7.78)$$

Replace the ideal PSR with the first approximation made in (7.66) to facilitate some cancellation in terms

$$\left| \sqrt{u^2 + r_0^2} + \frac{r_0 \Delta r}{\sqrt{u^2 + r_0^2}} - \sqrt{u^2 + r_0^2} - \Delta r \right| \leq \frac{\lambda}{8} \quad (7.79)$$

which reduces to

$$\left| \frac{r_0 \Delta r}{\sqrt{u^2 + r_0^2}} - \Delta r \right| \leq \frac{\lambda}{8} \quad (7.80)$$

Performing the usual binomial expansion on the radical in the denominator

$$\frac{\lambda}{8} \geq \left| \frac{r_0 \Delta r}{\sqrt{u^2 + r_0^2}} - \Delta r \right|$$

$$\begin{aligned}
&\geq \left| \frac{\Delta r}{\sqrt{\frac{u^2}{r_0^2} + 1}} - \Delta r \right| \\
&\geq \left| \Delta r \left(1 + \frac{u^2}{r_0^2} \right)^{-1/2} - \Delta r \right| \\
&\geq \left| \Delta r - \Delta r \frac{u^2}{2r_0^2} - \Delta r \right| \\
&\geq \left| -\Delta r \frac{u^2}{2r_0^2} \right|
\end{aligned} \tag{7.81}$$

We are interested in the magnitude of the error on the right side

$$\frac{\lambda}{8} \geq |\Delta r| \frac{u^2}{2r_0^2} \tag{7.82}$$

Defining u_{Max} as the maximum positive or negative along-track location of the platform during the collection (with aperture center set to $u = 0$), the restriction becomes

$$|\Delta r| \leq \frac{\lambda r_0^2}{4u_{Max}^2} \tag{7.83}$$

The DOF is twice this limit (one-half up-range plus one-half down-range)

$$DOF = 2|\Delta r| \leq \frac{\lambda r_0^2}{2u_{Max}^2} \tag{7.84}$$

Now, $D_{SAR} = 2u_{Max}$

$$DOF = 2 \frac{\lambda r_0^2}{D_{SAR}^2} \tag{7.85}$$

Finally, substituting the expression for cross-range resolution in (7.16) into (7.85) and ignoring any disparity between wavelength at the center frequency and away from the center frequency yields

$$DOF = 8 \frac{\Delta CR^2}{\lambda_c} \tag{7.86}$$

Given a fixed resolution and carrier frequency, (7.86) tells us that the depth of focus is independent of range to the swath. For the simulated example in Figure 7-37, where the cross-range resolution was 0.5 m and the wavelength was 0.0316 m, the DOF is equal to 63.3 m, just over 30 m up-range and down-range. At very fine resolutions and low frequencies the DOF grows so restrictive it becomes comparable to the down-range resolution, making this algorithm unwieldy.

7.4.3.3 Frequency-Domain PSR for RDA

A closed-form expression for the frequency-domain PSR given in (7.75) is needed to implement RDA as described already. The first step to solving (7.75) is to reindex time delay to place the closest point of approach of the hyperbola at time zero. This operation

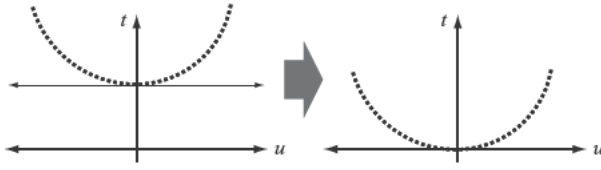


FIGURE 7-39 ■
Removal of time
offset from PSR.

is presented in a pictorial fashion in Figure 7-39. Removing the gross offset in time delay simply omits a large zero-padded interval in time; in frequency, it eliminates a steep linear phase modulation. There is no change to the final output image so long as the measured data is reindexed the same way.

Mathematically, this change to the PSR is

$$\delta \left(t - \frac{2}{c} \sqrt{u^2 + r_0^2} \right) \rightarrow \delta \left(t + \frac{2}{c} r_0 - \frac{2}{c} \sqrt{u^2 + r_0^2} \right) \quad (7.87)$$

After substituting (7.87) in for the reference PSR in (7.54), which in turn goes into (7.75), we have

$$H(k_u, \omega; r_0) \equiv \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \delta \left(t + \frac{2}{c} r_0 - \frac{2}{c} \sqrt{u^2 + r_0^2} \right) e^{-j\omega t} dt \right) e^{-jk_u u} du \quad (7.88)$$

where a preference for temporal integration before spatial integration is expressed. Equation (7.88) is solved using the principle of stationary phase (PSP) [5]

$$H(k_u, \omega; r_0) = A_1 \exp \left(j r_0 \left(\frac{\omega}{c/2} - \sqrt{\left(\frac{\omega}{c/2} \right)^2 - k_u^2} \right) \right) \quad (7.89)$$

$$A_1 = \sqrt{\frac{\pi c r_0}{j \omega}}$$

where the amplitude shading function, A_1 , is slowly varying and often ignored. The phase term is the driving force in RDA image formation; Figure 7-38a provides an example of this function.

7.4.3.4 RDA Approximations

A couple of approximations can be made to the quadratic phase term in the PSR in (7.89)

$$\psi_{PSR} = r_0 \left[\frac{\omega}{c/2} - \sqrt{\left(\frac{\omega}{c/2} \right)^2 - k_u^2} \right] \quad (7.90)$$

When valid, these approximations yield an image formation technique that, while still powerful, is much easier to implement computationally. It is this form of RDA that has seen widespread use [6,7]. As we will see, this approximation also makes it apparent why the algorithm appellation is *range-Doppler*.

It is often the case that slant-range spatial frequencies, $k_r = 2\omega/c$, are much higher than along-track spatial frequencies k_u . Slant-range spatial frequencies are provided by the carrier RF. For example, at C-band (5 GHz), the wavelength is 0.06 m, so the (two-way) spatial frequencies over the waveform bandwidth are on the order of 1 cycle per 0.03 m, or

33 cycles per m. For a 1 m cross-range resolution, (7.3) and (7.4) suggest the along-track sampling interval must be no larger than 0.5 m. If that corresponds to critical sampling then there will be signal content at the highest spatial frequencies of -1 cycle/m to $+1$ cycle/m. In this example there is a one-to-two orders-of-magnitude difference in spatial frequency. Assuming $2\omega c \gg k_u$, the radical inside the brackets in (7.90) can be replaced with the usual binomial approximation as

$$\begin{aligned} \frac{\omega}{c/2} - \sqrt{\left(\frac{\omega}{c/2}\right)^2 - k_u^2} &= \frac{\omega}{c/2} - \frac{\omega}{c/2} \sqrt{1 - \frac{k_u^2}{\left(\frac{\omega}{c/2}\right)^2}} \\ &\cong \frac{\omega}{c/2} - \frac{\omega}{c/2} + \frac{k_u^2}{2\left(\frac{\omega}{c/2}\right)} \\ &\cong \frac{ck_u^2}{4\omega} \end{aligned} \quad (7.91)$$

So, we denote the first approximation to the PSR by a single apostrophe

$$\psi_{PSR} \cong \psi'_{PSR} = \frac{cr_0 k_u^2}{4\omega} \quad (7.92)$$

Comparing (7.92) to (7.31), we see this approximation to the PSR in RDA is equivalent to the RMC term in DBS-AD-RMC, an interesting bridge over the gap between RDA and DBS.

The second approximation comes from making a narrowband assumption; that is, the waveform bandwidth is much less than the carrier frequency. Define a center frequency ω_0 and deviation from the carrier $\Delta\omega$ such that $\omega = \omega_0 + \Delta\omega$. Then (7.92) becomes

$$\psi'_{PSR} = \frac{cr_0 k_u^2}{4(\omega_0 + \Delta\omega)} \quad (7.93)$$

By assuming $\Delta\omega \ll \omega_0$ and applying the binomial approximation, the denominator in (7.93) reduces to

$$\begin{aligned} (\omega_0 + \Delta\omega)^{-1} &= \frac{1}{\omega_0} \left(1 + \frac{\Delta\omega}{\omega_0}\right)^{-1} \\ &\cong \frac{1}{\omega_0} \left(1 - \frac{\Delta\omega}{\omega_0}\right) \\ &\cong \frac{1}{\omega_0} - \frac{\Delta\omega}{\omega_0^2} \end{aligned} \quad (7.94)$$

We substitute the narrowband formulation into (7.93) and denote this second approximation by a double apostrophe:

$$\psi'_{PSR} \cong \psi''_{PSR} = \frac{cr_0 k_u^2}{4} \left[\frac{1}{\omega_0} - \frac{\Delta\omega}{\omega_0^2} \right] \quad (7.95)$$

This form of the PSR phase term

$$\psi''_{PSR} = \frac{cr_0 k_u^2}{4\omega_0} - \frac{cr_0 k_u^2}{4\omega_0^2} \Delta\omega \quad (7.96)$$

when substituted back into the PSR

$$H(k_u, \omega = \omega_0 + \Delta\omega; r_0) \approx \exp\left(j \frac{cr_0 k_u^2}{4\omega_0}\right) \exp\left(-j \frac{cr_0 k_u^2}{4\omega_0^2} \Delta\omega\right) \quad (7.97)$$

leads to two terms:

1. The first is a quadratic phase function in k_u . Because there is no dependence on RF, this modulation may be applied in either time or frequency.
2. The second is a linear phase modulation over baseband frequency $\Delta\omega$. Because multiplication by a linear phase is equivalent to a shift in the Fourier domain, the second term may be implemented as a time shift of the data, the degree of shift a function of spatial frequency.

Therefore, we may write this version of the PSR as

$$H(k_u, t; r_0, \omega_0) \approx \exp\left(j \frac{cr_0 k_u^2}{4\omega_0}\right) *_t \delta\left(t - \frac{cr_0 k_u^2}{4\omega_0^2}\right) \quad (7.98)$$

This form of the PSR is implemented by Fourier transforming the data $d(u, t)$ into the spatial-frequency domain $D(k_u, t)$, time-shifting each k_u bin by the second term in (7.98), phase modulating the data by the first term in (7.98), and inverse Fourier transforming from k_u back to u , which is now x in the image.

The time-delay shift in the second term in (7.98) is quadratic in k_u . It serves as an explicit straightening of the curved response of the PSR in (k_u, t) , an operation reminiscent of DBS-AD-RMC seen in Figure 7-26.

Note this version of RDA operates in the spatial-frequency–time-delay domain. As discussed earlier, spatial-frequency maps to *Doppler* frequency and time delay maps to slant *range*. This algorithm, then, operates in the *range-Doppler* domain, hence the moniker the range-Doppler algorithm.

This approximation to RDA offers two implementation advantages over the full matched filter development:

1. Fourier transforms from t to ω and back to t are avoided, while the time delays are easily realized using low-order interpolation filters.
2. Staying in the time domain makes it possible to vary the PSR with down-range. Indeed, the phase modulation can vary continuously with down-range. The time shift is more difficult since a finite length of range extent is needed for interpolation. Range variation on the time shift is commonly realized by processing the data in small but discrete blocks, or subswaths, wherein each subswath sees a different time-shift function. With down-range variation the PSR becomes

$$H(k_u, t; r, \omega_0) \approx \exp\left(j \frac{cr k_u^2}{4\omega_0}\right) *_t \delta\left(t - \frac{cr k_u^2}{4\omega_0^2}\right) \quad (7.99)$$

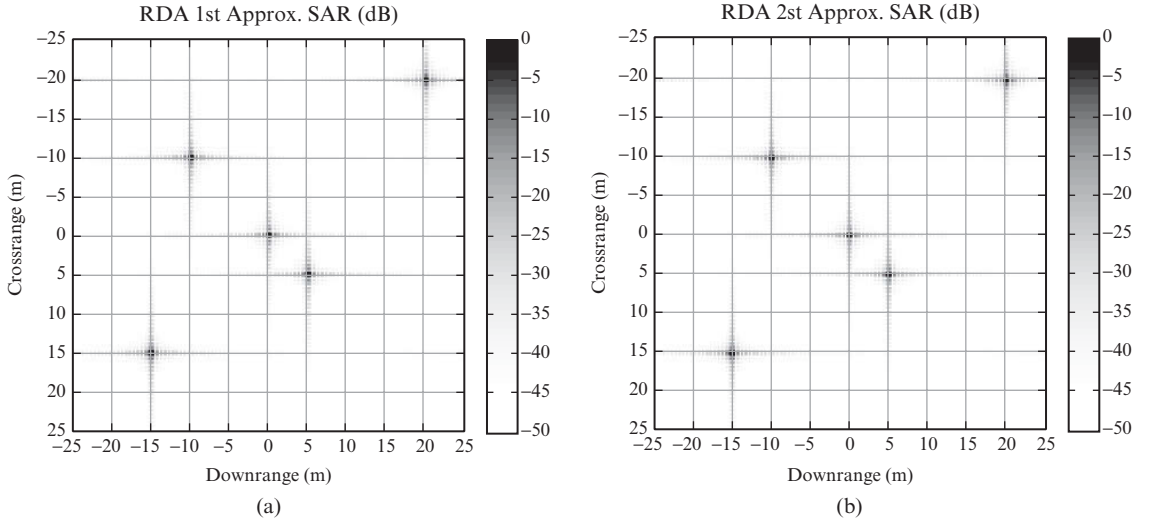


FIGURE 7-40 ■ Simulated RDA imagery for (a) first PSR approximation, and (b) second PSR approximation.

Relating the down-range at the closest point of approach to time delay yields

$$H(k_u, t; \omega_0) \approx \exp\left(j \frac{c^2 k_u^2}{8\omega_0} t\right) *_t \delta\left(t - \left(\frac{c^2 k_u^2}{8\omega_0^2} t\right)\right) \quad (7.100)$$

Equation (7.100) reflects the way RDA was historically implemented by the remote sensing community, with a continuously varying time shift approximated discretely over sub-swaths.

Figure 7-40 contains RDA images corresponding to that in Figure 7-37b and using the first PSR approximation in (7.92) in Figure 7-40a and the second in (7.98) in Figure 7-40b. The images in Figure 7-40 are indistinguishable in quality from the full-PSR image in Figure 7-37b.

7.4.3.5 The Modified Range-Doppler Algorithm

The binomial approximation was used to simplify the inverse of frequency in (7.93) to the difference between two frequency terms in (7.95). The binomial approximation is really the first two terms of the more general Taylor series expansion. Under many practical imaging conditions the third Taylor series expansion term becomes significant:

$$\begin{aligned} (\omega_0 + \Delta\omega)^{-1} &= \frac{1}{\omega_0} \left(1 + \frac{\Delta\omega}{\omega_0}\right)^{-1} \\ &\cong \frac{1}{\omega_0} \left(1 - \frac{\Delta\omega}{\omega_0} + \left(\frac{\Delta\omega}{\omega_0}\right)^2\right) \\ &\cong \frac{1}{\omega_0} - \frac{\Delta\omega}{\omega_0^2} + \frac{(\Delta\omega)^2}{\omega_0^3} \end{aligned} \quad (7.101)$$

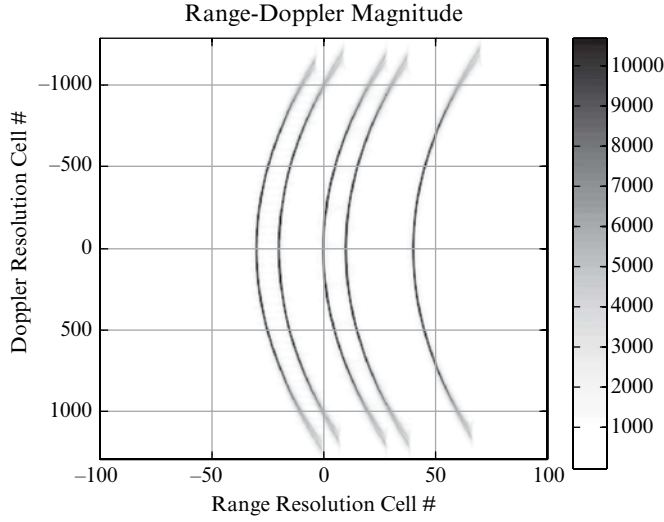


FIGURE 7-41 ■ PSRs in the spatial-frequency domain showing defocusing over time; image shows magnitude of complex voltage.

The PSR in (7.97) becomes

$$H(k_u, \omega = \omega_0 + \Delta\omega; r_0) \approx \exp\left(j \frac{cr_0 k_u^2}{4\omega_0}\right) \exp\left(-j \frac{cr_0 k_u^2}{4\omega_0^2} \Delta\omega\right) \exp\left(j \frac{cr_0 k_u^2}{4\omega_0^3} (\Delta\omega)^2\right) \quad (7.102)$$

This new term is quadratic in baseband frequency and therefore causes a defocusing of compressed range returns in (k_u, t) over t with increasing severity as the absolute value of k_u increases. Figure 7-41 shows five PSRs in the spatial-frequency domain for an L-band collection described in the next section, but with down-range to scene center increased to 25 km to amplify the artifacts of the quadratic frequency term in (7.102). Note the smearing over t (range) at the top and bottom of the image.

The defocusing caused by the quadratic term in (7.102) is a problem for RDA as represented by (7.98). After applying the k_u -dependent time delays to straighten the curvature of the PSRs each PSR should be a one-dimensional function over k_u at a constant t , but the quadratic term results in two-dimensional support. Compensation for this effect is achieved by multiplying the data by the conjugate of the quadratic term in the (k_u, ω) domain.

It would be inefficient to go from time to frequency and back to time again just to compensate for this additional term. However, pulse compression is usually performed through a conjugate multiplication in the frequency domain. That is, instead of correlating the time-domain data with the transmitted pulse waveform, the data and waveform are Fourier transformed from time to frequency and conjugate-multiplied together, and the product inverse Fourier transformed back time. One more Fourier transform from along-track to spatial frequency and we are in the (k_u, t) domain and ready to implement the RDA technique in (7.100). Now, pulse compression usually occurs in the slow-time (along-track position) domain but instead can be performed equivalently in the spatial-frequency domain, since along-track spatial frequency is orthogonal to time frequency. So, pulse compression may be implemented by first Fourier transforming the data from along-track to spatial frequency, then from time to frequency, multiplying by the conjugate of the frequency-domain form of the waveform, and finally inverse Fourier transforming

from frequency back to time. The modified range-Doppler algorithm alters the waveform matched filter with a k_u -dependent quadratic term designed to compensate for range defocusing. This adjustment to the pulse compression filter is known as secondary range compression (SRC) [8].

The key to modified RDA, then, is an alteration to the pulse compression procedure. To summarize, the technique is implemented as follows:

1. Two-dimensional Fourier transform the data from (u, t) to (k_u, ω) .
2. Generate the pulse compression matched filter by Fourier transforming the temporal waveform to ω .
3. Implement SRC by modifying the one-dimensional, ω -only matched filter to a two-dimensional, k_u -dependent filter using the third term in (7.102).
4. Multiply the data by this modified filter and inverse Fourier transform back to the (k_u, t) domain. The data are now pulse compressed, and the quadratic phase term in (7.101) is corrected.
5. Proceed with RDA as given in (7.100).

7.4.3.6 Notes on the Range-Doppler Algorithm

The RDA family is far superior to the DBS family, DBS-AD-RMC included. RDA takes full advantage of the hyperbolic PSR and benefits from computationally efficient matched filtering through frequency-domain multiplication. Unlike DBS, RDA is suitable for stripmap SAR because it is able to form an image over an arbitrarily long collection, neglecting memory constraints.

The Achilles' heel of full-PSR RDA is the variation in the PSR function over down-range. Recall RDA employs just one PSR, but many collection geometries can have a very limited DOF. An L-band simulation having the following parameter values serves to illustrate this point:

1. Carrier frequency = 1.5 GHz
2. RF bandwidth = 150 MHz
3. Integration angle = 5.4°
4. Resolution = 1.0 m (both dimensions)
5. Range to scene center = 4.1 km
6. Scene size = 100 m by 100 m

Diagrams of the collection geometry and the location of point targets are provided in Figure 7-42.

Figure 7-43a shows the RDA image generated with the PSR in (7.90) for the geometry in Figure 7-40, while Figure 7-43b is a similar image for a modified laydown of targets. Targets at down-range equal to zero, the range corresponding to the PSR used in RDA, focus well. Targets up-range and down-range become increasingly defocused.

The second approximation to RDA in (7.100) allows some compensation for this variation in range. However, several approximations were made to the matched filter to get to this form, and in practice the range variation is only partially compensated through block processing subswaths. We desire an image formation technique that uses the full PSR in (7.89) and accommodates the range variation of that PSR seamlessly. The range migration algorithm is that technique.

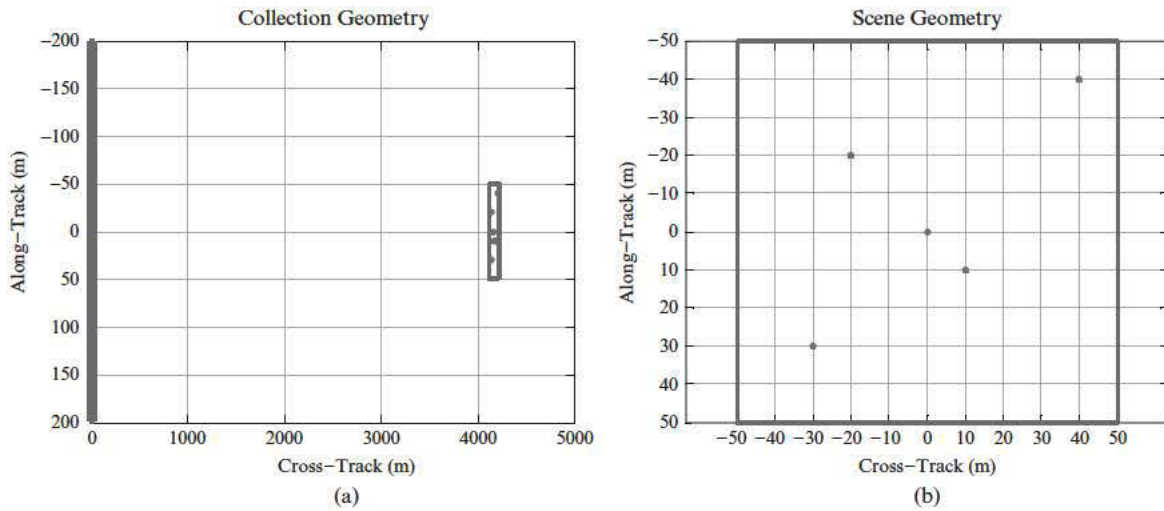


FIGURE 7-42 ■ Simulated RDA geometry emphasizing (a) the overall collection geometry, and (b) point scatterer locations in the scene of interest.

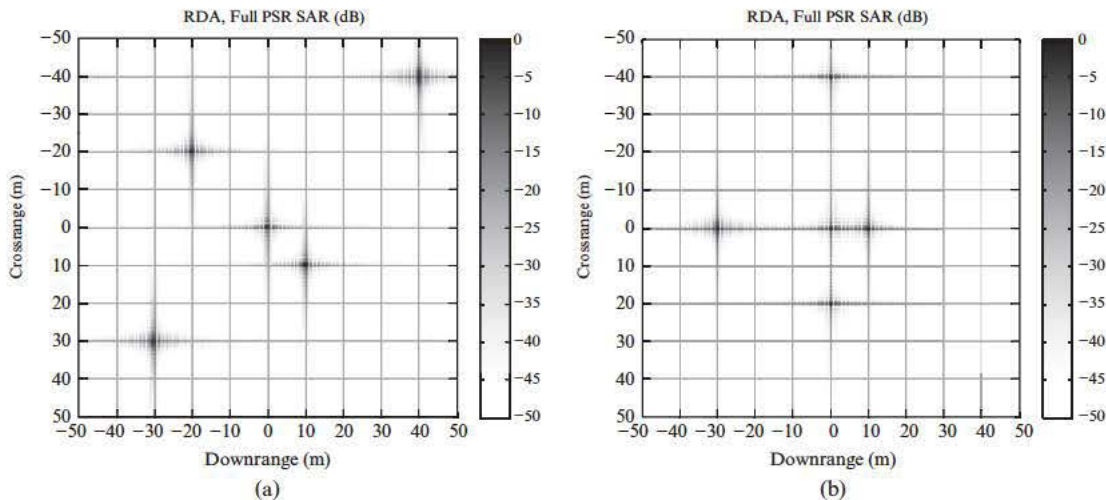


FIGURE 7-43 ■ Simulated RDA imagery using full-PSR for (a) point targets in one configuration, and (b) point targets in another configuration.

7.5 | RANGE MIGRATION ALGORITHM

The RMA was first developed within the seismic imaging community where it is known as the “omega-k” (or “ Ω -K” or “ ω -k”) algorithm because its key operations take place in the two-dimensional frequency domain [9]. Within a few years it was adopted by the stripmap SAR community as the range migration algorithm, perhaps because the technique fully accounts for the range migration of scatterers.

RMA is the culmination of our progression through image formers in this chapter. It is the algorithm of choice for challenging stripmap applications, where it has supplanted all varieties of DBS, RSA, and RDA. RMA fully exploits the PSR, making no approximations

in accounting for the down-range variation in the PSR, and benefits from a 2-D FFT efficiency.

7.5.1 RDA Revisited

RMA is effectively a modified implementation of RDA, so it is constructive to review some of the features of RDA. While a far more capable image formation approach than any variation on DBS, RDA suffers from a finite DOF, the consequence of a filter matched to just one PSR corresponding to a reference range. There are three ways to cope with this drawback:

1. Constrain the swath depth, the down-range extent of the image, to be less than the DOF. This restriction becomes unreasonable under imaging conditions having very small DOFs.
2. Partition the down-range swath into subswaths and image each subswath with RDA employing a PSR tailored to that down-range interval. Still, cross-range focus decreases in quality as one moves up-range or down-range with respect to the reference range in each subswath. In addition, each subswath must be smaller than the DOF; under challenging conditions the DOF may be on the order of the down-range resolution so that a separate RDA run is required for each down-range bin. Then this procedure is tantamount to the inefficient RSA.
3. By operating in the range-Doppler (time-delay and spatial-frequency) domain, RDA may be realized as a spatially variant interpolation and phase modulation. While this provides a mechanism for compensating for the down-range variation in the PSR, several approximations had to be made to the PSR to get to this implementation. In addition, making frequent changes to the range (time-delay) resampling function is awkward.

RMA starts with the full RDA implementation (no approximation to the PSR) and remedies its shortcoming by adding a resampling stage in the two-dimensional frequency, (k_u, ω) , domain prior to the 2-D inverse fast Fourier transform (IFFT) that produces the final image [10]. This Stolt interpolation takes data matched to just one PSR (at the reference range) and generates data matched at all down-ranges. In other words, Stolt interpolation yields an infinite DOF; each down-range bin is as well focused in cross-range as any other [11]. The resulting image has the quality of an RSA image (where we explicitly built a different PSR for each down-range bin) but generated with RDA 2-D FFT efficiency.

7.5.2 Coordinate Systems and Frequencies

In this section we change notation somewhat to be consistent with legacy RMA papers and book chapters.

Figure 7-44a summarizes the RMA coordinate system. We have been using u and x to denote cross-range (along-track), dedicating the former to the platform and the latter to scatterers, and will continue to do so. For down-range (cross-track) we have been using r for both the scene and the final image. With RMA we will use y for this dimension in the final image, so the image pixels are located as (x, y) . Finally, slant range, equal to the square root of the sum of the squares of along-track and cross-track differences, has been given as R , with time delay, t , equal to R divided by $c/2$.

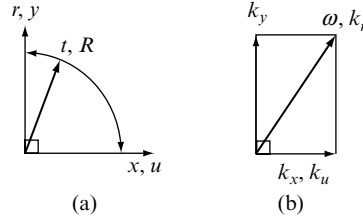


FIGURE 7-44 ■
RMA development
(a) coordinate
systems, and
(b) frequencies.

Figure 7-44b depicts the temporal and spatial frequencies of interest. The Fourier dual to u and x are k_u and k_x , respectively. Because u and x index into the same dimension these spatial frequency are equal, that is,

$$k_x = k_u \quad (7.103)$$

Consequently, a spatial frequency in the data, k_u , is equal to the same spatial frequency in the scene and image, k_x .

The Fourier dual to y is k_y ; thus, we will use k_y to denote the spatial frequency for the down-range dimension of the SAR image.

The Fourier dual to t is ω , which is equal to $2\pi f$, where f is RF in cycles/second. Alternatively, we may express this dimension as a spatial frequency k_r , where $k_r = 2\omega/c$ in radians/meter. Note that, in this development, k_r is not the Fourier dual of r . Finally, it can be shown that the spatial frequencies are related in the following way:

$$k_r^2 = k_x^2 + k_y^2 \quad (7.104)$$

7.5.3 Stolt Interpolation and RMA

Stolt interpolation is motivated by the following observations:

- Input data is in the (k_u, ω) domain.
- If the output is in the (k_x, k_y) domain, a 2-D IFFT is sufficient to produce an (x, y) image.
- Therefore, a transformation is needed for $(k_u, \omega) \rightarrow (k_x, k_y)$.

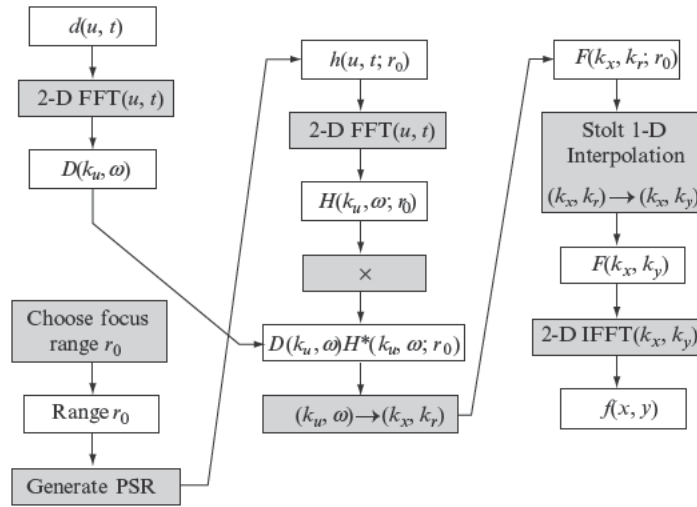
Equation (7.103) indicates the transformation on the first dimension is trivial. Stolt interpolation is concerned with the transformation along the second dimension. Equation (7.104) provides the form of this transformation

$$\begin{aligned} k_y &= \sqrt{k_r^2 - k_x^2} \\ &= \sqrt{\omega^2 \left(\frac{4}{c^2} \right) - k_x^2} \end{aligned} \quad (7.105)$$

Stolt interpolation, then, is a one-dimensional resampling of the frequency data. The along-track spatial frequencies are unaltered. Uniformly spaced samples in ω are interpolated using (7.105) to yield uniformly spaced samples in k_y .

Figure 7-45 provides a flow diagram for RMA that is equivalent to that shown in Figure 7-35 for RDA except for the renaming of frequencies, $(k_u, \omega) \rightarrow (k_x, k_r)$, and Stolt interpolation to (k_x, k_y) .

FIGURE 7-45 ■
Signal flow diagram
for the RMA.



7.5.4 Simulated RMA Examples

Let us revisit the L-band simulation example presented at the end of the RDA section. The image in Figure 7-46a is the same RDA image shown in Figure 7-43a, and Figure 7-46b is the corresponding RMA image. The only difference between these images is the additional of Stolt interpolation in Figure 7-46b, which serves to focus all scatterers, regardless of their down-range location.

Figure 7-47 depicts the geometry for a much more challenging imaging situation:

- Carrier frequency = 600 GHz
- RF bandwidth = 750 MHz
- Integration angle = 51°

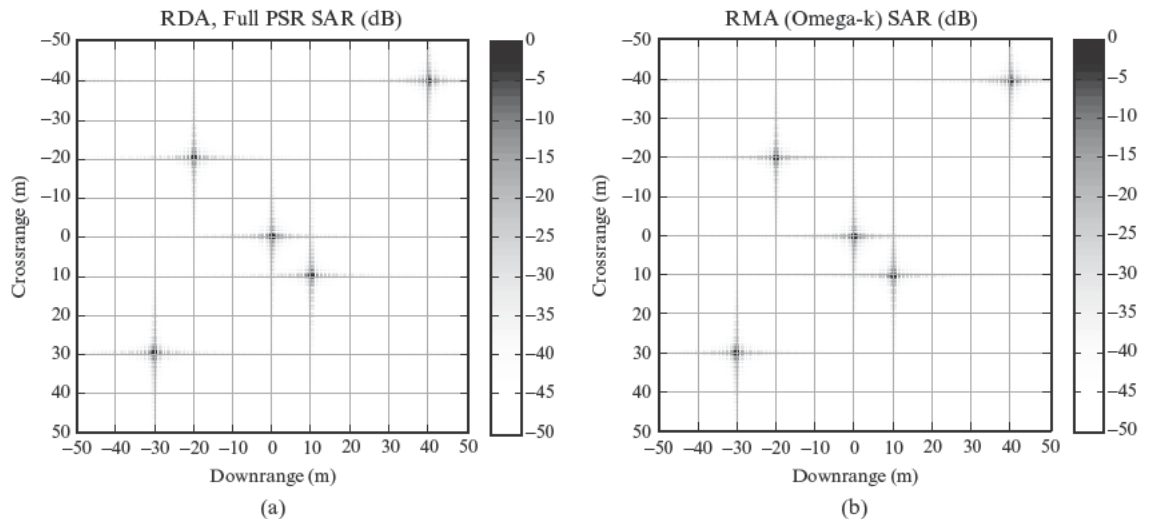


FIGURE 7-46 ■ Simulated L-band imagery using full-PSR and (a) RDA, and (b) RDA with Stolt interpolation (RMA).

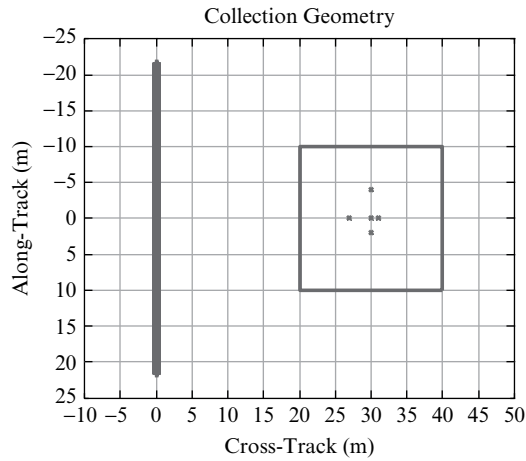


FIGURE 7-47 ■ Simulated UHF geometry.

- Resolution = 0.2 m (both dimensions)
- Range to scene center = 30 m
- Scene size = 10 m by 20 m

This same ultra high frequency (UHF) example was used in [1] to demonstrate back-projection image formation.

The RDA image in Figure 7-48a is well focused for the three scatterers corresponding to the PSR reference range in the middle of the swath. However, one does not need to move up-range or down-range very far before cross-range defocusing becomes unacceptable. In contrast, all five scatterers are well focused in the RMA image in Figure 7-48b, which compares favorably to the backprojection image in [1].

Figure 7-49 shows the data in the 2-D frequency domain, with k_u replaced by k_x and rescaled to k_r . The triangular support of the data is typical for wideband collections and is a consequence of the higher Doppler frequencies (larger values of k_x) supported by higher

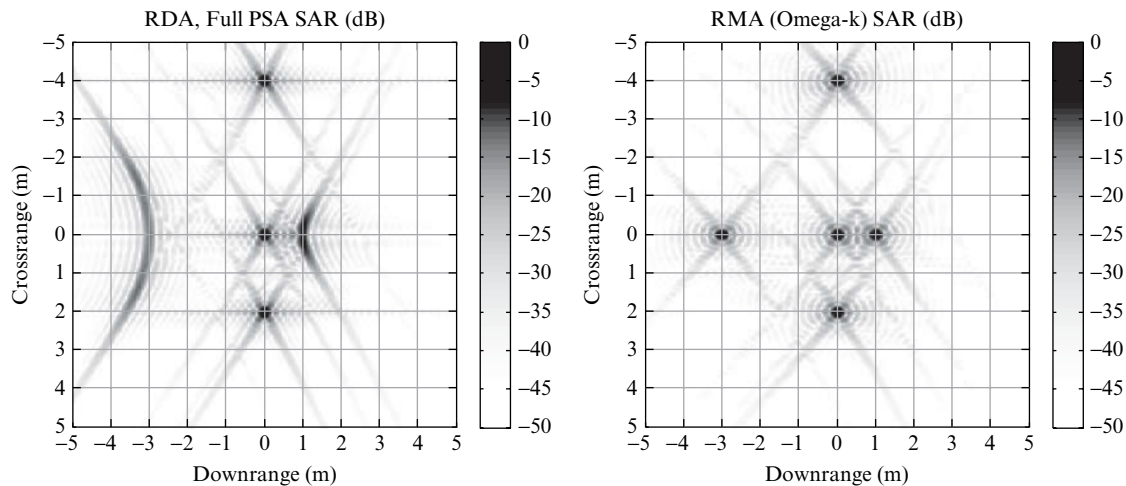


FIGURE 7-48 ■ Simulated UHF imagery using full-PSR and (a) RDA, and (b) RDA with Stolt interpolation (RMA).

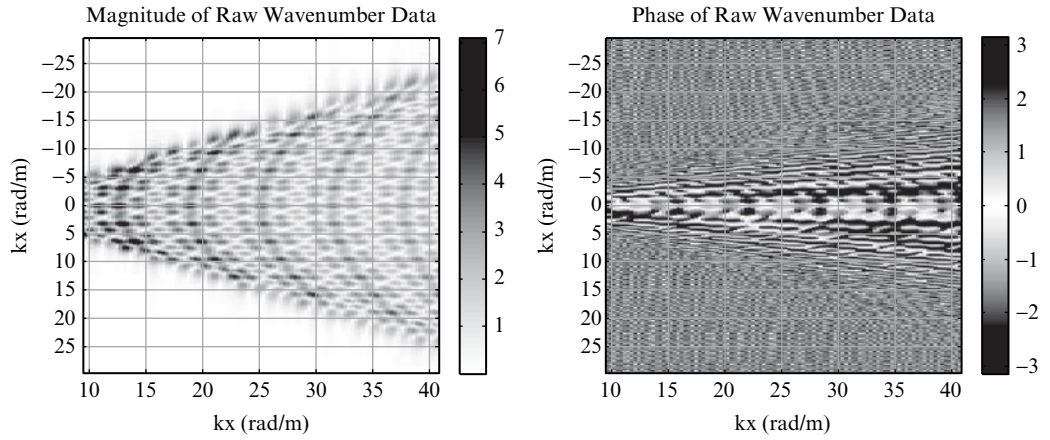


FIGURE 7-49 ■ Simulated UHF data in 2-D frequency domain (a) amplitude, and (b) phase.

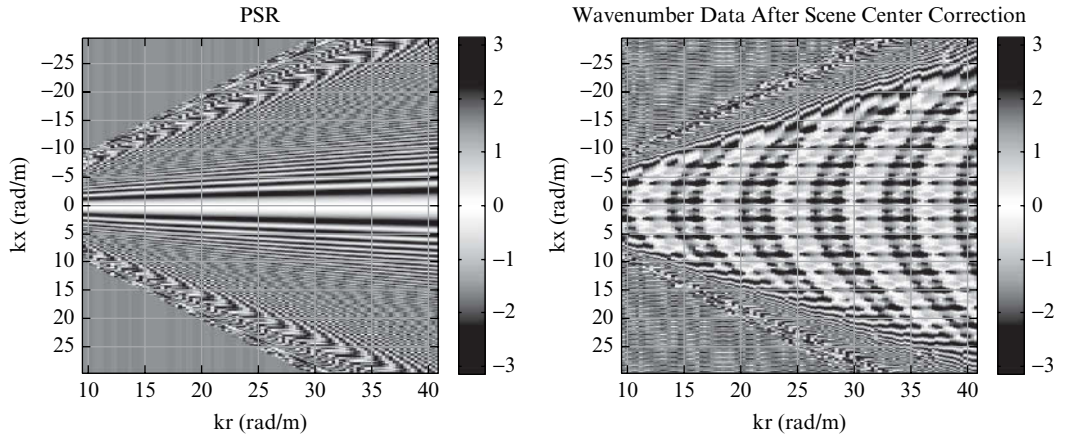


FIGURE 7-50 ■ Simulated UHF collection with phase of (a) PSR for matched filtering, and (b) data after application of PSR.

RF (larger values of k_r). The $k_x - k_r$ slope of the edge of this triangle is determined by the angle of integration; for stripmap acquisitions, by the antenna beamwidth. Figure 7-50a contains the PSR phase, and Figure 7-50b the phase of the data after application of the PSR matched filter. Figure 7-51 shows the data amplitude and phase after Stolt interpolation; compare Figures 49a to 51a and 50b to 51b. Notice the interpolation did not move data across the k_x rows; data is shifted and resampled along each k_x row only. We expect a collection of point targets to generate a periodic structure in the 2-D frequency domain; Figure 7-49a and 50b do not possess this property, while 50a and 50b do. The data in this (k_x, k_y) domain has the support of a slice of an annulus, where the inner radius (at left) is set by the lowest RF, the outer radius (at right) by the highest RF, and the angle limits at the top and bottom are set by the angle limits of integration.

The mechanics of Stolt interpolation are revealed in Figures 7-52, 7-53, and 7-54. The plots in Figure 7-52a show the locations of data measurements as a function of (k_x, k_r)

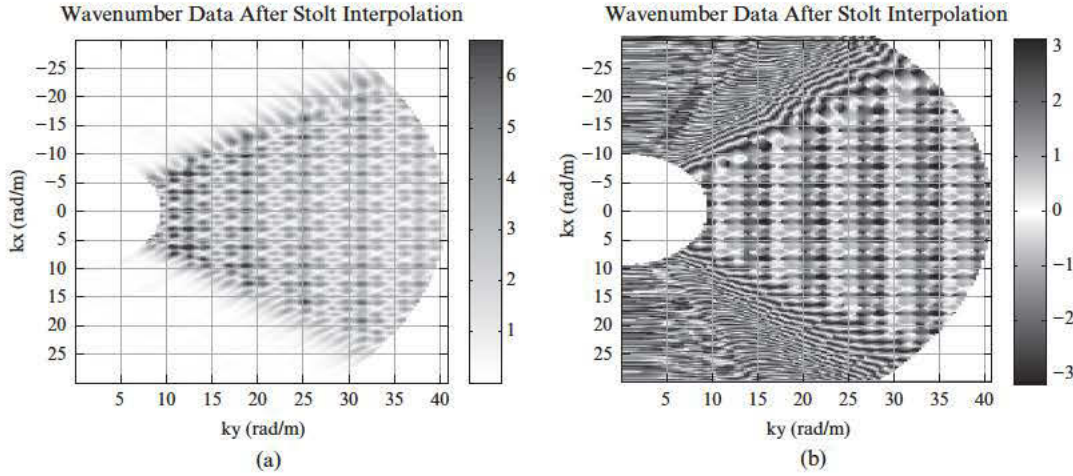


FIGURE 7-51 ■ Simulated UHF data in 2-D frequency domain after PSR matched filter and Stolt interpolation (a) amplitude, and (b) phase.

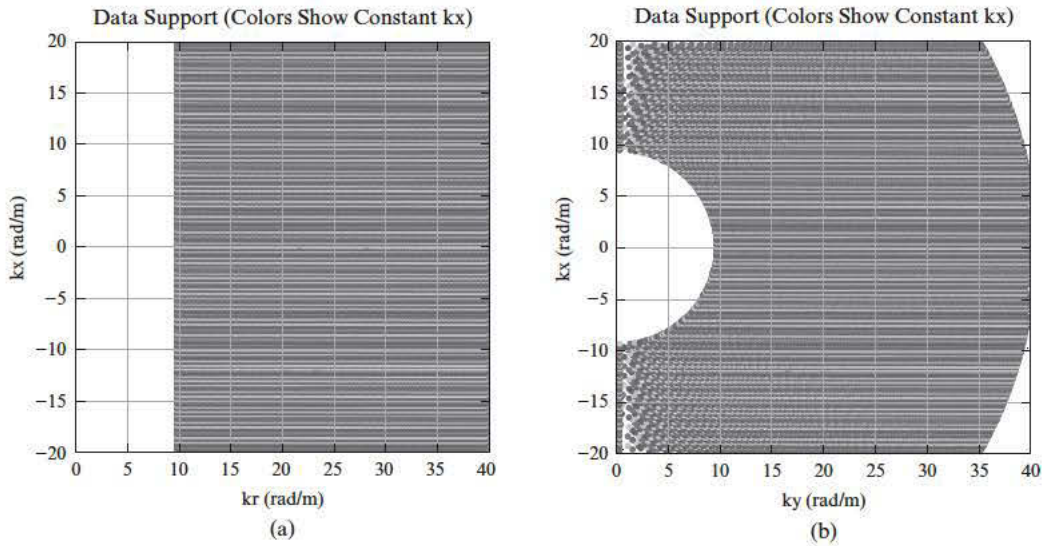


FIGURE 7-52 ■ UHF data measurement locations in (a) (k_x, k_r) , and (b) (k_x, k_y) .

using colored dots, where all dots of the same shade of gray (in a column) have the same k_r value. This data fall onto a rectangular grid. The corresponding plot in Figure 7-52b shows the (k_x, k_y) locations of data measurements, where (7.105) was used to generate the k_y values. The measured data do not fall onto a rectangular grid in this domain. Figure 7-53 is an expanded view of a subset of the frequency extent shown in Figure 7-52. As the arrows suggest, the new k_y values do not modify the k_x locations; k_y is a shift and resampling of the k_r data along each k_r . This shift is more significant for large k_x magnitudes. Note that this is not simply a bulk shift but is a k_r -dependent interpolation, as (7.105) suggests. To perform the resampling we calculate the locations of the measurements in (k_x, k_y) in Figure 7-54a (same as Figure 7-53a), establish a rectangular grid of desired (k_x, k_y)

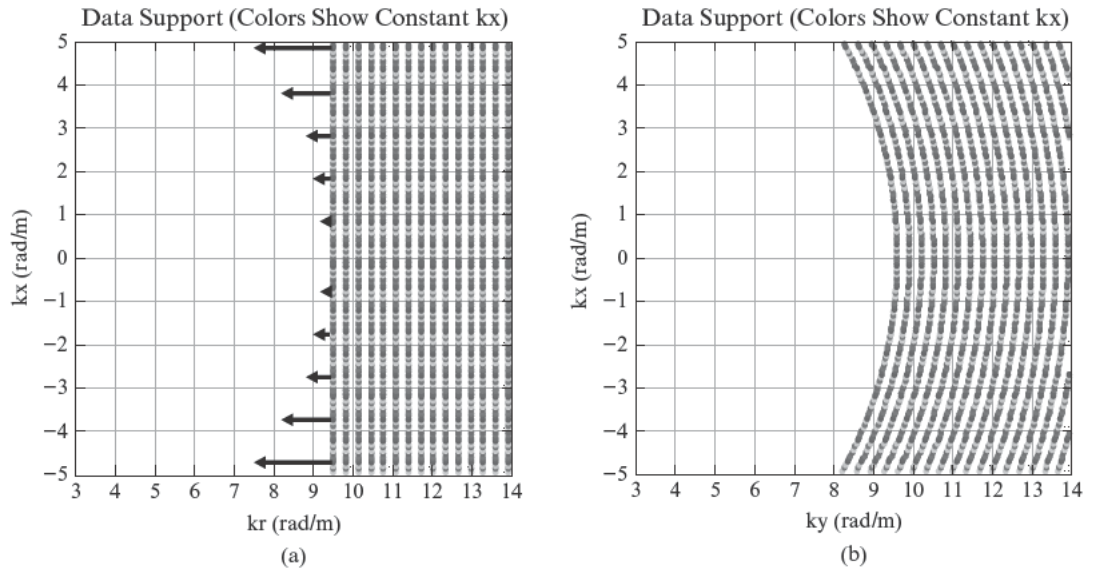


FIGURE 7-53 ■ Subset of UHF data measurement locations in (a) (k_x, k_r) , and (b) (k_x, k_y) . The arrows suggest the shift and resampling of the k_r data along each k_r .

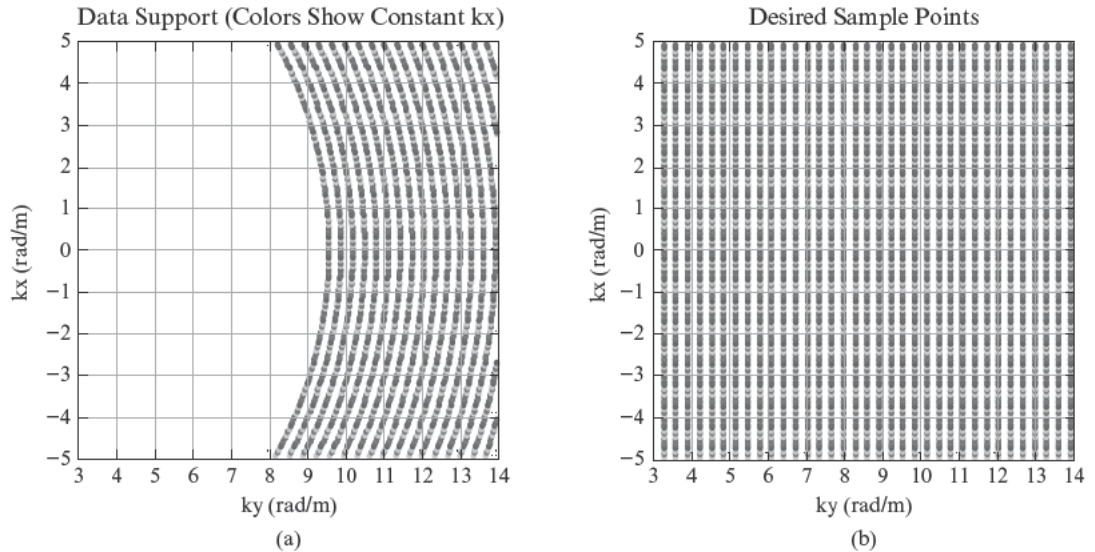


FIGURE 7-54 ■ Simulated UHF data in (k_x, k_y) with (a) measured locations, and (b) target locations.

locations, shown in Figure 7-54b, and interpolate. Zeros are applied to target locations outside the measured data region of support.

We have yet to derive or justify Stolt interpolation. Like the two-dimensional frequency representation of the PSR (7.89) used in RDA and RMA, Stolt interpolation is found through an analysis of matched filter imaging using the PSP. An excellent mathematical derivation of both the form of the PSR matched filter and Stolt interpolation using the PSP is found in [11].

7.5.5 Stolt Approximation

In the previous section the point was made that the Stolt interpolation is not just a bulk shift of the data from k_r to k_y for each set of k_x values. But perhaps Stolt interpolation could be approximated in this way.

To justify this approximation mathematically, begin with the exact mapping from k_r to k_y in (7.105) and assume that $k_x \ll k_r$. (Recall this was the first approximation made to develop the time-domain implementation of RDA.) Then we can employ the binomial expansion to get rid of the radical

$$\begin{aligned} k_y &= k_r \left(1 - \frac{k_x^2}{k_r^2} \right)^{1/2} \\ &\cong k_r - \frac{k_x^2}{2k_r} \end{aligned} \quad (7.106)$$

Define a center spatial frequency for k_r as k_{r0} . For a sufficiently small fractional bandwidth (the second approximation made to develop the time-domain implementation of RDA), the k_r in the second term of (7.106) may be replaced with a fixed center frequency k_{r0}

$$k_y \approx k_r - \frac{k_x^2}{2k_{r0}} \quad (7.107)$$

For a given value of k_x in (7.107) the second term is a constant, and mapping for k_r to k_y consists of a bulk shift of the data along k_r . This constant delay may be realized with low-order sinc interpolators. Alternatively, exploiting the relationship that a constant delay is a linear phase modulation in the Fourier domain, one could inverse Fourier transform the data from k_r to r , apply a k_x -dependent linear phase over r to effect the shift in (7.107), and then Fourier transform back to k_r , which is now k_y . Since the next step to range compression with an inverse Fourier transform from k_y to y , we might as well omit both the transform back and the inverse transform forward. That is, stay in the (k_x, r) domain, which is actually the (k_x, y) domain after the linear phase multiply, and finish with a Fourier transform from (k_x, y) to (x, y) . This sequence is depicted in the flow diagram in Figure 7-55.

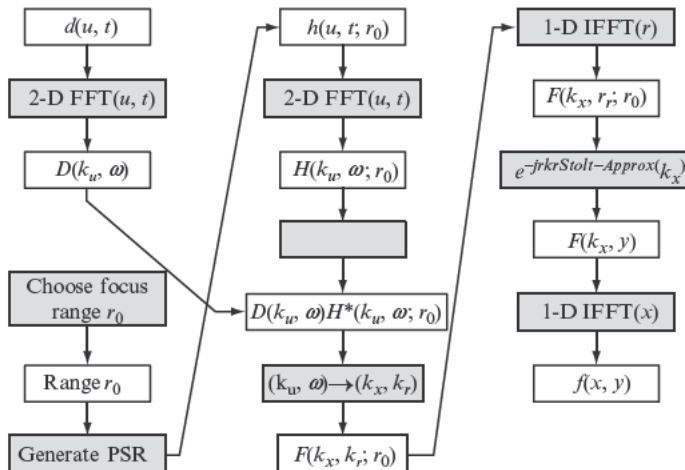


FIGURE 7-55 ■
Signal flow diagram
for the RMA with
approximation to the
Stolt interpolation.

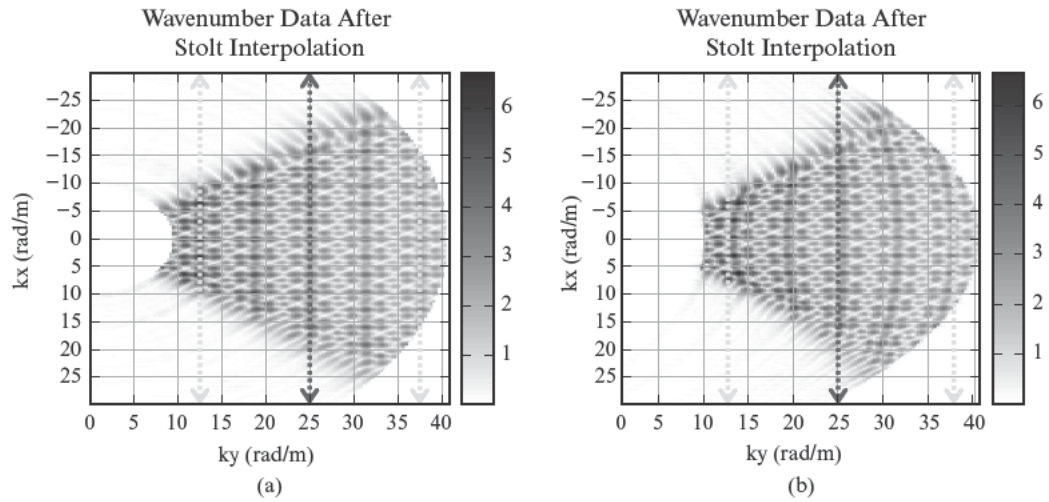


FIGURE 7-56 ■ Simulated UHF data in (k_x, k_y) after (a) exact Stolt interpolation, and (b) approximate Stolt interpolation.

The performance of the RMA using the exact and approximated Stolt interpolations on the simulated UHF data is shown in Figure 7-56. Dashed arrows are provided as reference. The match along the middle arrow at $k_y = 25$ is excellent, as this wavenumber corresponds to the k_{r0} value used in the approximation in (7.107). (A close examination shows that the match is not perfect along this arrow, due to the first approximation made in (7.106).) In examining the left and right arrows, however, residual bowing can be observed in Figure 7-56b that does not appear in Figure 7-56a. This approximation to the Stolt interpolation has its limits.

Figure 7-57 contains RMA images using the Stolt interpolation in Figure 7-55 for UHF and L-band data. The UHF image in Figure 7-57a is better than the RDA image in Figure 7-48a but not as good as the RMA image generated with the exact Stolt interpolation in Figure 7-48b. The approximation to Stolt interpolation was not appropriate for this

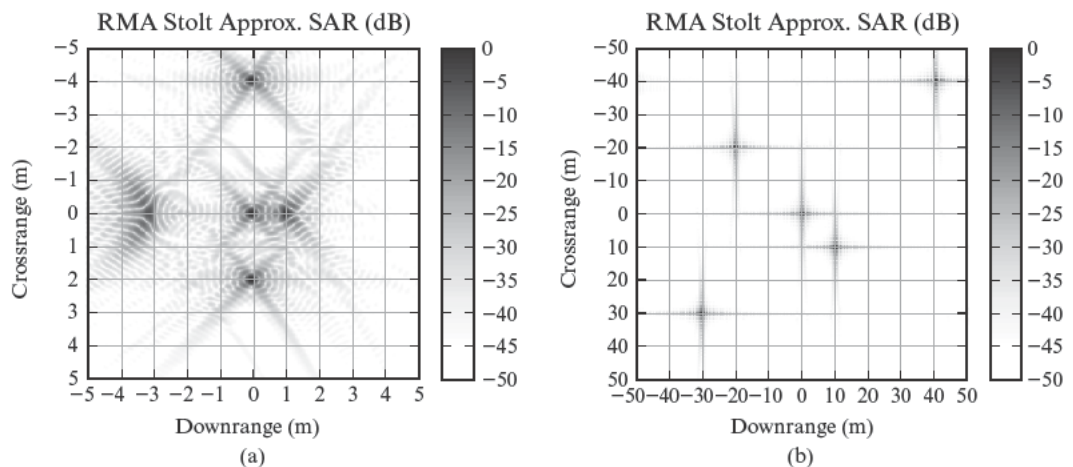


FIGURE 7-57 ■ RMA SAR imagery using Stolt approximation for simulated (a) UHF data, and (b) L-band data.

example. The scatterers in the L-band image in Figure 7-57b are focused and compare well with those in the exact Stolt image in Figure 7-46b except for slightly higher down-range sidelobes on the scatterers at the upper left and lower right. The Stolt approximation was appropriate here.

7.5.6 Chirp Scaling Algorithm

Another option for approximating the Stolt interpolation in RMA is found in the chirp scaling algorithm (CSA). CSA may be understood by revisiting RDA [12]. Recall two kinds of RDA were developed:

1. Matched filtering with the exact PSR from a reference down-range via a phase multiplication in the two-dimensional frequency (k_u, ω) domain. Cross-range focusing is error-free at the reference range but gets increasing worse as one moves away from the reference range.
2. Matched filtering with an approximation to the PSR, implemented as a phase modulation and time shifts in the (k_u, t) domain. Focusing is not perfect anywhere (because approximations were made to the PSR to get to this implementation), but the spatial variation of the PSR over down-range may be accounted for to a certain extent by allowing the phase and shift functions to vary over time delay t .

An intriguing amalgamation of these two procedures can be imagined as follows:

1. Match filter with the exact PSR via a phase multiplication in (k_u, ω) , as in (1) in the previous list.
2. Transform to (k_u, t) and compensate for residual errors in the responses of scatterers displaced from the reference range through differential time shifts. By differential we mean with respect to the reference range. In other words, no time shift is applied at the reference range but shifts become nonzero up-range and down-range from the reference.

This procedure realizes error-free focusing at the reference range, like RDA using the exact PSR, and improves focusing away from the reference range, like the approximation to RDA.

The typical CSA implementation realizes these stages in a reversed order:

1. Transform to (k_u, t) and apply differential time shifts to give PSRs up-range and down-range forms more closely matched to the PSR at the reference range.
2. Transform to (k_u, ω) and apply the matched filter for the PSR at the reference range.

Then the SAR image is formed with a 2-D IFFT as in Figure 7-35.

The differential time shifts in the first step need to be defined. Bulk time shifts t_{Shift} for the approximation to RDA are found in the final term of (7.100)

$$t_{Shift} = \frac{c^2 k_u^2}{8\omega_0^2} t \quad (7.108)$$

For the reference range r_0 the time delay is $t_0 = r_0/(c/2)$, and the resulting time shift is

$$t_{Shift}(t_0) = \frac{c^2 k_u^2}{8\omega_0^2} t_0 \quad (7.109)$$

The differential time shift with respect to the reference t_{Dif} is then

$$\begin{aligned} t_{Dif}(t - t_0) &\equiv t_{Shift}(t) - t_{Shift}(t_0) \\ &= \frac{c^2 k_u^2}{8\omega_0^2}(t - t_0) \end{aligned} \quad (7.110)$$

Properties of the differential time shift function in (7.110) include:

1. Zero shift at the reference time t_0 , as expected; matched filtering with the reference PSR in (k_u, ω) fully accounts for the target responses at this time delay.
2. Time shift increases with the magnitude of k_u ; in fact, it is quadratic with k_u .
3. Time shift increases linearly as time delay moves away from the reference time. The differential time shift is positive as we go down-range ($t > t_0$) and negative as we move up-range ($t < t_0$).

An important point to keep in mind is that (7.110) provides the form of the PSR, not how to implement the matched filter. Recall the matched filter in the frequency domain takes the form of the conjugate of the PSR. Conjugation reverses the sign of time delays. So, for image formation, time shifts are applied to the data with a sign opposite to that as given in (7.110). The effect is to straighten the curvature predicted in the measured data by (7.110). For imaging, we want to implement negative time delays (pull returns toward time equal zero) for data down-range from the reference and do the opposite for data up-range.

Differential time shifts for a nonzero k_u are depicted in stylized form in Figure 7-58a. The middle response is at the reference range/time and does not experience a time shift. Other responses are shifted in time toward t_0 as shown, with increasing shifts the farther the response from the reference time. These shifts might be realized with a spatially variant interpolation filter, an awkward implementation.

Now assume the radar returns have not yet been pulse compressed, and the radar employs an LFM waveform. In fact, LFM is the norm for imaging applications. Further assume that the radar digitizes and records the full waveform bandwidth; it does not perform so-called stretch processing, also known as deramp-on-receive. Figure 7-58b shows the recorded data as a function of frequency for the five scatterers depicted in Figure 7-58a. As expected, the data consist of the coherent superposition of multiple chirp signals. The compensation in (7.110) would still have to be implemented as time shifts on the data. Indeed, this processing is more difficult to realize in Figure 7-58b because the uncompressed returns from the scatterers now overlap in time.

LFM waveforms possess a property known as range-Doppler coupling. Modulating an LFM return with a constant frequency (e.g., a Doppler offset) effects a range displacement after pulse compression. In Figure 7-58c, frequency offsets are applied to the LFM returns instead of time delays. These offsets are chosen to correspond to the time delays depicted in Figure 7-58b. Except for artifacts generated by the beginning and end of each LFM, the offset signals in Figure 7-58c would, after pulse compression, have the desired time shifts shown in Figures 58a and 58b.

Figure 7-58d shows the tones required to cause the frequency offsets in Figure 7-58c. Negative frequency modulations are applied to returns before the reference and positive frequencies to time delays beyond the reference.

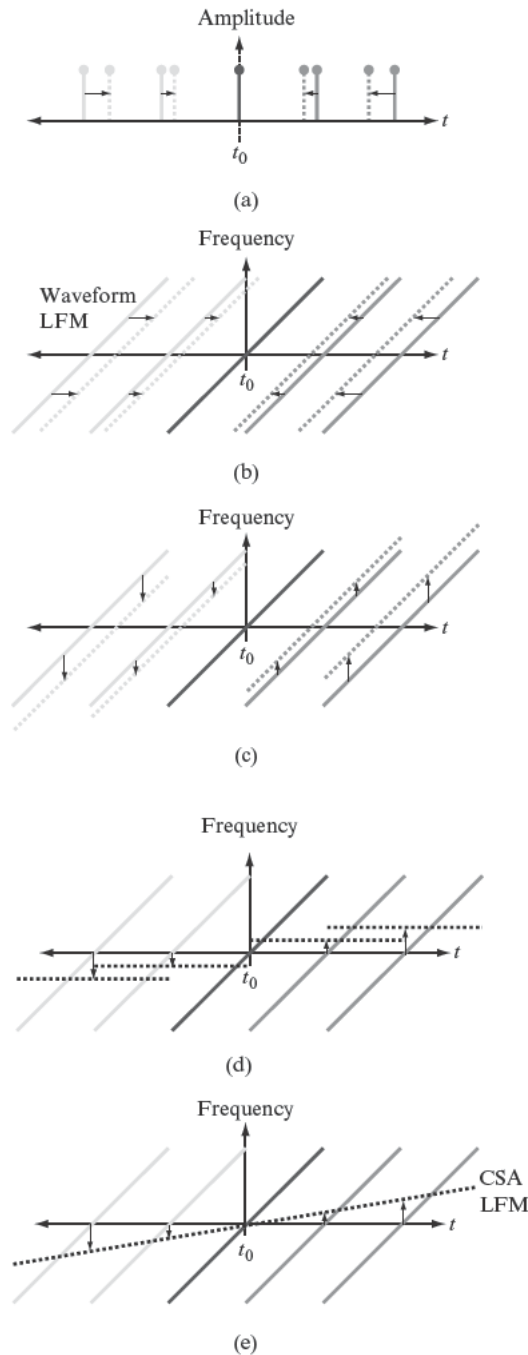


FIGURE 7-58 ■ Development of CSA modulation function, (a) time shifts to compressed returns versus time delay, (b) same time shifts applied to uncompressed LFM returns, (c) time shifts realized through range-Doppler coupling, (d) modulation signals used to induce time shifts, (e) LFM signal used to induce time shifts.

Finally, Figure 7-58e shows a shallow LFM approximation to the discretized stepped tones in Figure 7-58d. By applying this CSA LFM to the chirped data the time shifts shown in Figure 7-58a are realized without interpolation of the data.

In summary, the time-varying time shifts in (7.110) required to correct for the variation in the PSR response with range can be achieved by modulating LFM waveform returns

with a very gradual CSA LFM. CSA, then, is implemented as follows [13]:

1. Full-bandwidth LFM waveform returns are collected and not compressed to form $d(u, t)$. If recorded data have already been deramped or pulse compressed they must be rechirped before application to CSA.
2. The recorded data are Fourier transformed to $D(k_u, t)$.
3. LFM's are applied over t , a different LFM for each k_u , to effect the differential time shifts in (7.110).
4. The data are Fourier transformed to $D(k_u, \omega)$ or, equivalently, $D(k_x, k_y)$. However, (3), compensation for down-range variation in the PSR, is the approximation to Stolt interpolation. Therefore, the data are already in the $D(k_x, k_y)$ domain.
5. Pulse compression is performed with a frequency-domain phase multiplication. A second phase multiplication is applied to realize the RDA matched filter. These operations may be implemented as a single aggregate phase multiplication. Modulation functions have heretofore been defined in the ω domain (for pulse compression) and (k_u, ω) domains (for RDA) and so must be modified to be consistent with the (k_x, k_y) data.
6. A 2D-IFFT produces the output image.

Several nuances and additional compensation steps have been neglected in this development of CSA. The aim here was to introduce CSA in the context of RDA and RMA, highlighting LFM multiplication of uncompressed LFM waveform returns as a means to approximate Stolt interpolation. A detailed development of CSA is found in [14].

7.6 | OPERATIONAL CONSIDERATIONS

Issues impacting stripmap system performance [15] and implementation are explored in this section.

7.6.1 PRF Lower Limits

In [1] a minimum along-track sampling rate was established and was replicated in this chapter in (7.4). This requirement for a sampling interval less than or equal to one-fourth the physical aperture size was motivated by a desire to unambiguously sample returns over the null-to-null beamwidth. Here we replicate this development but reference sampling in time, not distance, and frequency in Doppler, not wavenumber, to prepare for a more extensive examination of PRF bounds.

At broadside, the nominal geometry for a stripmap acquisition, the Doppler bandwidth, B_d , for a cone angle extent, $\Delta\theta$, centered on broadside is approximately

$$B_d = \frac{2v}{\lambda} \Delta\theta \quad (7.111)$$

where v is the platform ground speed. If the peak-to-null beamwidth is denoted as θ_R , the null-to-null Doppler bandwidth, B_{2R} , is

$$B_{2R} = \frac{4v}{\lambda} \theta_R \quad (7.112)$$

A good approximation for the peak-to-null beamwidth is given by

$$\theta_R \approx \frac{\lambda}{D_{az}} \quad (7.113)$$

where D_{az} is the azimuth (horizontal) length of the radar antenna. The null-to-null bandwidth is then

$$B_{2R} = \frac{4v}{D_{az}} \quad (7.114)$$

The PRF must be greater than or equal to this bandwidth to capture the beamwidth without aliasing

$$PRF \geq \frac{4v}{D_{az}} \quad (7.115)$$

Under the condition of equality in (7.115) the Doppler bandwidth available for SAR processing, B_{SAR} , is equal to B_{2R} . This situation is shown in Figure 7-59a with the null-to-null Doppler extent of clutter depicted by the elliptical spectral support.

PRFs lower than the limit in (7.115) are acceptable so long as the unaliased Doppler spectrum is low-pass filtered and the consequent spoiling of cross-range resolution is acceptable. For example, in Figure 7-59b the PRF has been reduced by 3/4 below the limit in (7.115). B_{SAR} is reduced by a factor of two with respect to Figure 7-59a but is still reasonable. Indeed, this bandwidth is consistent with the 3 dB beamwidth of the antenna that was used to derive cross-range resolution relationships like (7.3). Figure 7-59b corresponds to an along-track sampling interval of one-third the aperture size, in contrast to the one-fourth limit given in (7.4).

The PRF limit in (7.115) may be violated by as much as a factor of two, leaving an infinitesimally small bandwidth for SAR. This situation is depicted in Figure 7-59c.

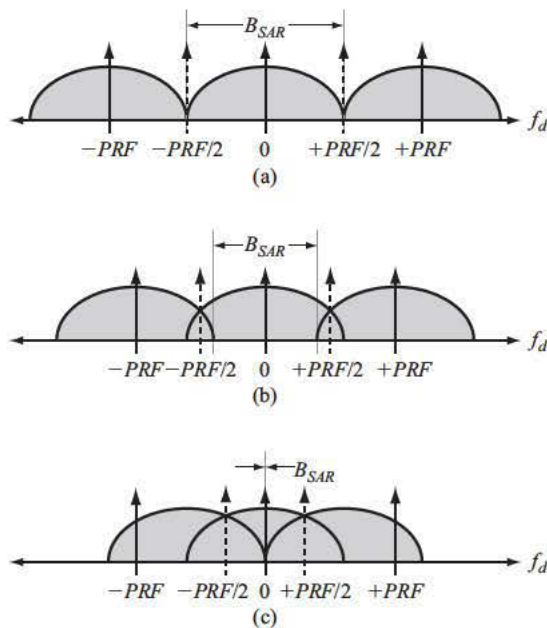


FIGURE 7-59 ■
Doppler spectrum of mainbeam clutter with PRF, (a) matched to the null-to-null extent, (b) PRF decreased by a factor of 4/3, and (c) PRF decreased by a factor of 2.

Let us represent this relaxation of the azimuth null-to-null PRF constraint with an azimuth sampling factor, K_{AZ} , that takes on a value between 1 and 2. The PRF constraint in (7.115) can be written

$$PRF \geq \frac{4v}{K_{AZ} D_{az}} \quad (7.116)$$

The impact on B_{SAR} is easily found as

$$B_{SAR} = B_{SAR}^0 \left(\frac{2}{K_{AZ}} - 1 \right) \quad (7.117)$$

where B_{SAR}^0 is the achievable Doppler bandwidth for unambiguous sampling of the main-beam. Reduction of the SAR Doppler bandwidth increases (spoils) stripmap resolution by the same amount, so that (7.3) may be written

$$\Delta CR(\text{stripmap}) \geq \frac{D_{az}}{2} \left(\frac{K_{AZ}}{2 - K_{AZ}} \right) \quad (7.118)$$

If K_{AZ} equals 1, the full Doppler bandwidth of the mainbeam is available for imaging and the finest stripmap resolution is possible. If K_{AZ} equals 2, the PRF limit is relaxed by a factor of 2 but there is no unaliased bandwidth for SAR processing and cross-range resolution is infinite.

The PRF restriction in (7.116) establishes a limit on swath depth and area coverage rate. If the PRF is too high, returns from multiple pulses will arrive at the radar simultaneously. These range-ambiguous returns are avoided if the swath depth is such that returns from the up-range edge on the current pulse arrive at the radar after returns from the down-range edge on the previous pulse:

$$T_{Swath} \leq \frac{1}{PRF} \quad (7.119)$$

The swath extent in slant range is $cT_{Swath}/2$, and this interval is projected onto the ground, L_{Swath} , by the cosine of the local grazing angle, θ_g ,

$$L_{Swath} = \frac{cT_{Swath}}{2 \cos \theta_g} \quad (7.120)$$

Combining (7.119) and (7.120) establishes a limit on the ground swath depth

$$L_{Swath} \leq \frac{c}{2 \cos \theta_g PRF} \quad (7.121)$$

Combining (7.116) and (7.121) yields

$$L_{Swath} \leq 8K_{AZ} \left(\frac{c}{v} \right) \frac{D_{az}}{\cos \theta_g} \quad (7.122)$$

Deeper swaths are afforded by slower platform speeds and larger antennas.

Area coverage rate (ACR) is equal to the swath depth times the velocity of the platform

$$ACR = vL_{Swath} \quad (7.123)$$

Combining (7.122) and (7.123) yields a limit on ACR

$$ACR \leq 8K_{AZ} \frac{cD_{az}}{\cos \theta_g} \quad (7.124)$$

The ultimate limit on ACR is set by the length of the physical aperture.

7.6.2 PRF Upper Limits

It is often the case that the ground swath is limited by elevation beamwidth of the antenna. If the peak-to-null elevation beamwidth is denoted ϕ_R , the null-to-null beamwidth is $2\phi_R$. At a slant range of R , the width across the elevation beamwidth is $2R\phi_R$. This extent is projected onto the ground by the sine of the local grazing angle. The resulting beam-limited swath depth, $L_{Swath,BW}$, is therefore

$$L_{Swath,BW} = \frac{2R\phi_R}{\sin \theta_g} \quad (7.125)$$

A good approximation to the peak-to-null beamwidth is

$$\phi_R \approx \frac{\lambda}{D_{el}} \quad (7.126)$$

where D_{el} is the elevation (vertical) length of the radar antenna. Combining (7.125) and (7.126) yields

$$L_{Swath,BW} = \frac{2R\lambda}{D_{el} \sin \theta_g} \quad (7.127)$$

If we apply the swath-induced PRF limit in (7.121) to (7.127), an antenna-induced limit on PRF results

$$PRF \leq \frac{cD_{el}}{4R\lambda} \tan \theta_g \quad (7.128)$$

This PRF limit ensures the swath extent on the ground bounded by the elevation mainbeam provided by an antenna of height D_{el} is imaged without aliasing in range.

A higher PRF than the limit suggested in (7.128) is permissible if only a fraction of the illuminated ground swath is processed. In the extreme, an infinitesimally small swath extent in the center of the illuminated swath may be processed. Then, range-ambiguous returns are avoided if the return from this patch on the current pulse arrives just after the return from the previous pulse on the down-range edge of the illuminated extent and just before the return from the next pulse on the up-range edge of the illuminated extent. Because ambiguities are defined by half the illuminated extent and not the full extent assumed in (7.128), the PRF limit in (7.128) may be relaxed by a factor of two.

Let us represent this relaxation of the elevation null-to-null PRF constraint with an elevation sampling factor, K_{EL} , that takes on a value between 1 and 2. The PRF constraint in (7.128) can be written as

$$PRF \leq K_{EL} \frac{cD_{el}}{4R\lambda} \tan \theta_g \quad (7.129)$$

The resulting unaliased processed swath extent, $L_{Processed}$, can be easily found as

$$L_{Processed} = L_{Processed}^0 \left(\frac{2}{K_{EL}} - 1 \right) \quad (7.130)$$

where $L_{Processed}^0$ is the elevation beamwidth limit on the processed swath in (7.127). Note that (7.130) has a form similar to that seen in (7.117). Setting $L_{Processed}^0$ to $L_{Swath,BW}$ yields

$$L_{Processed} = \frac{2R\lambda}{D_{el} \sin \theta_g} \left(\frac{2}{K_{EL}} - 1 \right) \quad (7.131)$$

Then the ACR in (7.123) can be modified with the processed swath extent in (7.131) as

$$\begin{aligned} ACR &= vL_{Processed} \\ &= \frac{2R\lambda v}{D_{el} \sin \theta_g} \left(\frac{2}{K_{EL}} - 1 \right) \end{aligned} \quad (7.132)$$

7.6.3 Antenna Area and System Utility

Now combine the lower PRF limit due to azimuth considerations in (7.116) with the upper PRF limit due to elevation considerations in (7.129):

$$\frac{4v}{K_{AZ} D_{az}} \leq PRF \leq K_{EL} \frac{c D_{el}}{4R\lambda} \tan \theta_g \quad (7.133)$$

Allowing the left and right sides in (7.133) to meet produces a constraint on the antenna size

$$D_{az} D_{el} \geq 16 \frac{v}{c} \left(\frac{1}{K_{AZ} K_{EL}} \right) \frac{R\lambda}{\tan \theta_g} \quad (7.134)$$

If this antenna area cannot be met, a PRF cannot be found that does not suffer from some level of range aliasing over the processed swath or Doppler aliasing over the processed bandwidth. The azimuth and elevation factors allow the antenna area requirement to be relaxed with a consequent reduction in swath depth or spoiling of cross-range resolution. Note that slant range and platform speed figure prominently in (7.134). The relatively slow speeds and short ranges in airborne collections means the minimum antenna area is usually easy to meet. Space-based applications, with their associated higher speeds and longer ranges, tend to motivate employment of much larger apertures.

We can define a SAR figure of merit for operational utility U_{SAR} as the ratio of ACR and resolution:

$$U_{SAR} = \frac{ACR}{\Delta CR(stripmap)} \quad (7.135)$$

Utility grows with increasing ACR and finer cross-range resolution. Substituting the expression for resolution in (7.118) and ACR in (7.132) into (7.135) yields

$$U_{SAR} = \frac{4R\lambda v}{(D_{az} D_{el}) \sin \theta_g} \left(\frac{2}{K_{EL}} - 1 \right) \left(\frac{2}{K_{AZ}} - 1 \right) \quad (7.136)$$

Utility is inversely proportional to antenna area. Hence, a sufficiently high antenna area is needed to provide for a PRF that avoids mainbeam range and Doppler aliasing, but further increases in antenna area serve to reduce utility. Indeed, (7.134) and (7.136) allow us to

TABLE 7-2 ■ System Design Parameters and Performance

Parameter	Airborne FOPEN	Space-Based SAR	SAS
c (m/s)	300 million	300 million	1,500
Grazing angle	30°	30°	30°
Utility limit (m/s)	86.6 million	86.6 million	430
Slant range (m)	20,000	1 million	100
Speed (m/s)	150	7,000	1.5
Frequency (Hz)	450 million	5 billion	30,000
Wavelength (m)	0.67	0.06	0.05
Antenna area limit (m²)	0.185	39	0.14
Antenna length (m)	1.0	15	0.5
Resolution (m)	0.5	7.5	0.25
Antenna height (m)	0.5	3	0.3
Antenna area (m ²)	0.5	45	0.15
Beam-limited swath (m)	105,000	800,000	67
ACR (m/s)	16 million	560 million	100
Realized utility (m/s)	32 million	75 million	400
PRF lower limit (Hz)	600	1,867	12
PRF upper limit (Hz)	1,625	2,165	13

put a limit on utility:

$$U_{SAR} \leq \frac{c}{4 \cos \theta_g} (2 - K_{EL}) (2 - K_{AZ}) \quad (7.137)$$

Utility is limited primarily by the propagation speed of light.

Once an antenna area is selected that meets the limit in (7.134) but provides acceptable utility as given in (7.136), antenna length and height can be traded against one another to emphasize fine stripmap cross-range resolution, as given in (7.118), or a high area coverage rate, as given in (7.132) with an associated swath depth. Finally, when antenna dimensions are set with ACR and resolution, (7.133) is used to settle on a good operating PRF.

Table 7-2 provides design parameters and predicted performance for three notional system designs: An airborne UHF FOPEN SAR, a C-band space-based SAR, and an underwater synthetic aperture sonar (SAS). Azimuth and elevation factors are set to 1.0. Bold rows denote measures of performance derived from parameters set in previous rows.

The formulas and examples in the section are meant to show trends and ignore curved-earth and large-angle effects. For example, the swath for the airborne FOPEN systems is just over 100 km, but the slant range to the swath center is only 20 km, clearly an unrealizable combination.

7.6.4 Slant Plane and Ground Plane

The PSR presented in (7.5) does not explicitly account for platform height or target/scene height. Fortunately, all of the stripmap image formation techniques we developed with this PSR are still applicable to nonzero platform and scatterer heights. To justify this assertion, let h denote the platform height and z account for the height of a scatterer in the imaged

scene. The PSR in (7.5) generalizes to

$$R(u, h; x, r_g, z) = \sqrt{(u - x)^2 + r_g^2 + (h - z)^2} \quad (7.138)$$

where r_g is the cross-track (down-range) distance from the platform to the scatterer at the closest point of approach ($u = x$) in the ground plane (platform and scatterer projected to zero height). Define a cross-track (down-range) distance r in a slant plane as

$$r \equiv \sqrt{r_g^2 + (h - z)^2} \quad (7.139)$$

where the slant plane is delineated by the trajectory of the platform and the scatterer. If the platform is flying straight level and the scatterer is stationary, then r_g, h , and z are all constant over the collection, and r will be a constant. The PSR then reduces to

$$R(u; x, r(r_g, h, z)) = \sqrt{(u - x)^2 + r^2} \quad (7.140)$$

which is same form in (7.5).

In short, the PSR remains hyperbolic if the platform or scatterers have nonzero height, so the resulting RMA image will still focus. (An RDA image will also be in focus, DOF limitations notwithstanding.) Indeed, any straight-line path taken by the platform, even climbing or diving, results in a hyperbolic PSR and, in turn, well-focused stripmap imagery.

A scatterer's down-range location in the image generated by RDA or RMA will be given by the slant plane, r , and not the ground plane, r_g . A mapping from the slant plane to the ground plane may be created using (7.139):

$$r_g = \sqrt{r^2 - (h - z)^2} \quad (7.141)$$

Note that the platform and scatterer heights (or their relative difference, at the least) must be known to perform this mapping. Nonzero heights leave along-track (cross-range) positions unaffected, so the mapping from slant plane to ground plane involves a one-dimensional resampling of the data as given by (7.141).

7.6.5 Other Imaging Approaches

In this chapter, the processing of stripmap measurements has emphasized *stripmap* image formation techniques, algorithms tailored to stripmap geometries and phenomenologies. However, so-called *spotlight* image formation techniques are a viable alternative under many circumstances. Spotlight algorithms operate by first designating a scene center location, sometimes known as the center reference point (CRP). Next, sufficient data along-track is recorded and sequestered to yield the desired cross-range resolution at the CRP and over a larger scene extent. Then, the requisite block of data is time delayed and phase corrected to constrain a return at the CRP to one range bin and inform it with a constant phase. This process is known as MCP, a procedure that is incompatible with stripmap image formation techniques like RDA and RMA. Finally, the motion-compensated data are applied to a tomographic image formation technique like the polar-formatting algorithm (PFA). MCP, PFA, and other aspects of spotlight data collection and imaging are described in detail in Chapter 6 in this volume.

To apply spotlight imaging to a stripmap collection, the desired output scene is partitioned along-track into multiple adjoining subscenes. Partitioning may also occur in

cross-track (down-range) as well, if a deep swath is desired. A CRP is designated for each subscene, and a subset of the measured data is sequestered for spotlight imaging of each subscene. The resulting set of spotlight images are stitched together to form the full stripmap image.

The spotlight images must be small enough to ensure discontinuities across image boundaries are minimized. Also, as discussed already, ample along-track data must be processed to meet resolution-driven integration angle requirements over an entire subscene. This combination of small subscenes and sizable data subsets means there is considerable overlap between the subsets applied to different subscenes. In other words, measurements on any given pulse are applied to the image formation of at least two subscenes. The application of each measurement to several spotlight images hints at some of the inefficiencies in this approach as compared to a dedicated stripmap algorithm.

There is a rich history of partitioning a long stripmap scene and imaging each of the subscenes with overlapping blocks of data. For example, this approach was used to overcome some of the limitations to SPECAN (DBS with AD) discussed at the end of Section 7.3.3. The step transform processing method [3] applies short, overlapped blocks of data to DBS-AD to generate overlapped, coarse-resolution subimages along-track. These subimages are registered along-track with respect to one another, and a second Fourier transform is applied through the subimages to realize the full, fine cross-range resolution potential of the collection.

Finally, stripmap data acquisitions may be applied to the backprojection image former. As discussed in [1], backprojection is the brute-force, space-time domain realization of the generalized matched filter in (7.50). Backprojection garners no computationally efficiency through FFT processing, so image formation may take much more time than RDA and RMA. On the other hand, backprojection accommodates nonlinear and nonuniformly spaced acquisitions, while RSA, RDA, RMA, and all DBS methods require linear, uniformly spaced sampling along-track. Therefore, backprojection may be the only image formation option given highly unstructured collection geometries. For example, data collected by BoomSAR, a ground-based low-frequency SAR system, suffered from severe motion errors generated by the swaying of the extended boom upon which the radar antenna array was mounted [16]. While the nominal vehicle motion was linear, large along-track and cross-track antenna motion precluded Fourier-based image formation, leaving backprojection as the only suitable imaging algorithm for this stripmap collection.

7.6.6 Squinted Operation

A stripmap collection was defined in the Introduction to this chapter by three attributes. Two of these properties, the platform moves in a straight line and the antenna beam patterns are not steered during the collection, will continue to hold. The third, radar antennas are pointed to broadside, will be relaxed in this section.

In the context of stripmap SAR, *broadside* means normal to the platform velocity vector defined with respect to the imaged ground swath. Operation with transmit and receive antenna beams pointed away from broadside is termed *squinted*. Squinted stripmap may occur for a number of reasons.

For instance transmit and receive beams may be mechanically or electronically steered (and fixed to an angle) either forward or aft of broadside. Small squint angles offer no significant advantage or disadvantage with respect to cross-range resolution or area coverage rate but serve to complicate image formation. In this situation the need to squint is often

driven by larger system and resource allocation considerations. For example, a SAR may be capable of both stripmap and spotlight modes and may be tasked with several such collections on a pass. A broadside stripmap collection would be preferred, but perhaps the system will be busy performing higher-priority spotlight collections at that time, relegating the stripmap task to earlier or later in the pass and therefore requiring squinted operation.

Directing beams to broadside or squinting them fore or aft does not affect the hyperbolic slant-range history and the form of the PSR; these are driven solely by the geometry. However, the beam patterns provide a shading to PSR. At broadside the portion of the PSR defined by the mainbeam is symmetric about the closest point of approach. Squinting the beam highlights the leading side of the PSR at the expense of the trailing side or vice versa. The resulting asymmetry is tantamount to a bulk change in average slant range to scatterers as they move through the beam. This linear migration of scatterers through range, so-called range walk, is the same for all scatterers at the same down-range location. Range walk is equivalently described in temporal frequencies as a Doppler offset or in along-track wavenumbers as a spatial-frequency offset.

The fix for squint-induced range walk depends on the image formation technique. Because the PSR remains hyperbolic, the filter used by RMA and RDA is still a match to the data; all that remains is to ensure the data and filter have the same support in the (k_u, ω) domain. However, the time shifts applied to the data in the spatial frequency domain in the approximation to RDA become difficult to implement when range walk is significant. It is prudent to remove range walk from the data prior to the Fourier transform from along-track to spatial frequency. Since walk is linear and the same for all scatterers, correction is a straightforward linear shift in time and linear multiply in phase over the along-track collection. To first order, all that remains after correction is a hyperbolic range curvature that is symmetric about the closest point of approach. The data are now ready for Fourier transformation to the spatial-frequency domain and application to the established time-domain RDA method. This same preprocessing of squinted data to remove a bulk, linear range walk and phase progression is also appropriate for the DBS family of techniques.

Squinted operation may also be a consequence of motion outside the platform. For example, winds aloft may change the effective motion of an airborne SAR. The apparent velocity of the platform from the imaged ground is the vector sum of the platform velocity with respect to the surrounding air and the velocity of the wind with respect to the ground. Similarly, planetary rotation will modify the collection geometry of an orbiting space-based SAR. From the point of view of a satellite, the rotating earth is manifested as a spatially variant vector flow field. However, for the typically small SAR beam footprint the earth's rotation can be approximated as a spatially invariant velocity vector. Then, like the airborne system in wind, the velocity of the platform with respect to the imaged ground is the vector sum of the orbital velocity and the rotational velocity.

Both airborne winds aloft and space-based planetary rotation may be modeled as modified platform velocity vectors with respect to the imaged swath. There are four important consequences to this model:

1. The apparent motion is still linear, so the hyperbolic PSR development is relevant to image formation.
2. In general, the apparent velocity vector has a modified magnitude, so the along-track sampling interval is, for a fixed PRF, larger or smaller or expected. (Or the PRF must be modified to maintain a desired along-track sampling interval.)

3. In general, the apparent velocity vector has a modified orientation, resulting in a rotation of the stripmap SAR image.
4. The vehicle is no longer aligned with the velocity vector; that is, crabbing occurs. If the radar antenna is oriented broadside to the platform, then it will be squinted with respect to the collection.

Once orientation and sampling are accounted for, the principle consequence of wind or planetary rotation is a squinted collection as described earlier for a steered antenna. Therefore, the approach to image formation is the same: application of the hyperbolic PSR for RMA and RDA, and linear time delays and phase modulations to remove bulk range walk for the time-domain implementation of RDA and for DBS.

7.7 | APPLICATIONS

As discussed in the Introduction, stripmap operation is a special case of a continuum of collection options. The distinguishing feature of stripmap is the radar antenna is fixed to broadside. Alternatively, the antenna beam could be mechanically or electronically steered during the collection. For example, the beam could be steered fore (starting toward the direction of travel) and allowed to drift aft in such a way as to continuously illuminate passing scenes of interest. This spotlight mode of operation facilitates increased integration angles and finer cross-range resolutions with the downside of sparse, interrupted coverage along-track with reduced ACR. Alternatively, the antenna could be continuously scanned in elevation or periodically scanned aft to fore along two or more depression angles. In this ScanSAR mode of operation the realized integration angle is less than that for stripmap; therefore, cross-range resolution is spoiled but a larger swath is surveyed, so ACR is increased.

Operating from ScanSAR through stripmap to spotlight allows resolution and ACR to be traded against one another in a continuous way. Why, then, is the special, zero-scan stripmap mode so common? There are three driving conditions:

1. The antenna cannot be scanned, leaving only stripmap as a mode option.
2. The antenna beamwidth is so wide that scanning would produce negligible changes in cross-range resolution and ACR.
3. Continuous swath coverage is desired along-track at the finest possible resolution.

A stripmap mode is often employed in remote sensing endeavors, especially from spaceborne platforms, and by FOPEN radars. Sensor and system parameters for one example each of these two applications appear in Table 7-3. A dash is used where values could be neither found nor derived.

7.7.1 Remote Sensing Applications

Remote sensing is the measurement and estimation of physical parameters of an object or phenomenon without physical contact. In modern parlance, remote sensing applies to the collection of measurements and retrieval (estimation) of parameters for the earth's surface from airborne or space-based sensors.

SAR technology has been very successfully applied to problems in remote sensing. Examples include [17]:

TABLE 7-3 ■ Example Stripmap Sensor and Platform Parameters

Parameter	RADARSAT-1	P-3 UWB SAR
Background		
Application	Remote sensing	FOPEN
Operators	Canadian Space Agency	ERIM/NAWC
Platform location	Space-based	Airborne
Year	1995	1994
Geometry		
Slant range (km)	1,000	6.24 km
Platform altitude (km)	798	3.1
Grazing angle (deg)	49	30
Platform velocity (m/s)	7,400	135
Platform ground speed (m/s)	6,600	135
Antenna		
Antenna size (m)	15 × 1.5	—
Antenna beamwidth (deg)	0.2	115–125 (frequency-dependent)
Illumination time (s)	0.5	25
Integration angle (deg)	0.2	32
Polarization	HH	HH,HV,VH,VV
Waveform		
Carrier frequency (GHz)	5.3 (C-band)	0.47 (UHF)
Waveform type	LFM	LFM
Waveform bandwidth (MHz)	30 (widest)	509
Pulse width (us)	42 or 43	—
Chirp rate (MHz/us)	0.72	—
PRF	1,300	—
Processing		
Waveform sampling	Direct	Stretch (deramp-on-receive)
Sampling rate (MHz)	32.2	—
Image formation	CSA	RMA
SAR performance		
Swath width (km)	50	0.9
Slant-range resolution (m)	5	0.33
Ground-range resolution (m)	8	0.4
Cross-range resolution (m)	8	0.66

1. Geology: Two-dimensional and three-dimensional cartography; change detection due to flooding, volcanic eruption, and earthquake activity; surface roughness and associated composition and weathering.
2. Ecology: Land cover/use classification; biomass measurement; flood mapping and delineation of wetlands.
3. Hydrology: Soil moisture and snow wetness measurements.
4. Oceanography: Characterization of surface waves, currents, and winds; monitoring of shipping and ship wakes.
5. Ice: Iceberg mass and motion; glacial ice topography and flow velocity.

Other uses include management of forest and agricultural resources and preparation for and recovery from natural disasters such as floods and tsunamis.

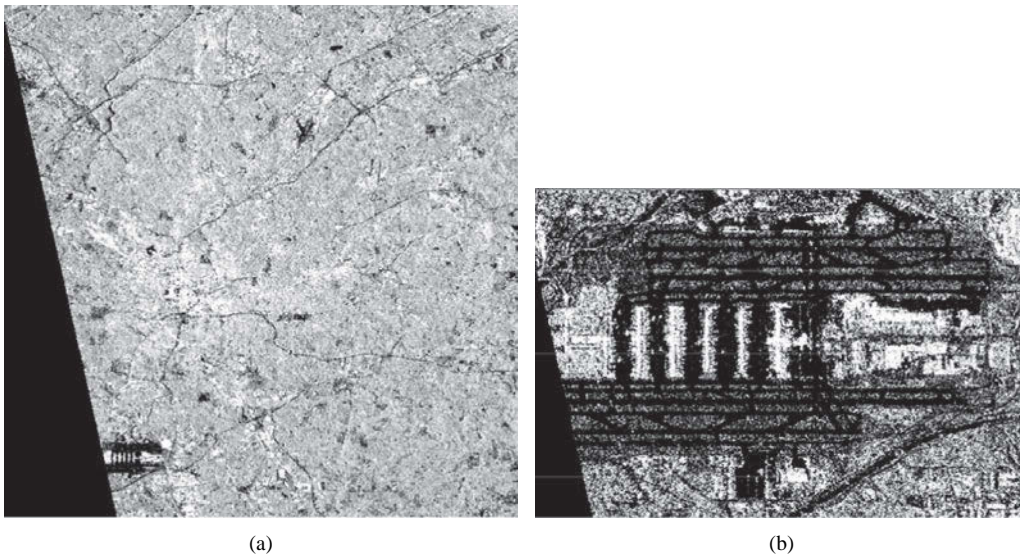


FIGURE 7-60 ■ Example RADARSAT-1 imagery (a) stripmap image of Atlanta, GA, and vicinity, and (b) expanded view of Atlanta's Hartsfield-Jackson International Airport. (RADARSAT-1 Data © Canadian Space Agency/ Agence spatiale canadienne ____1998. Received by the Canada Centre for Remote Sensing. Processed and distributed by MDA Geospatial Services Inc. All rights reserved.)

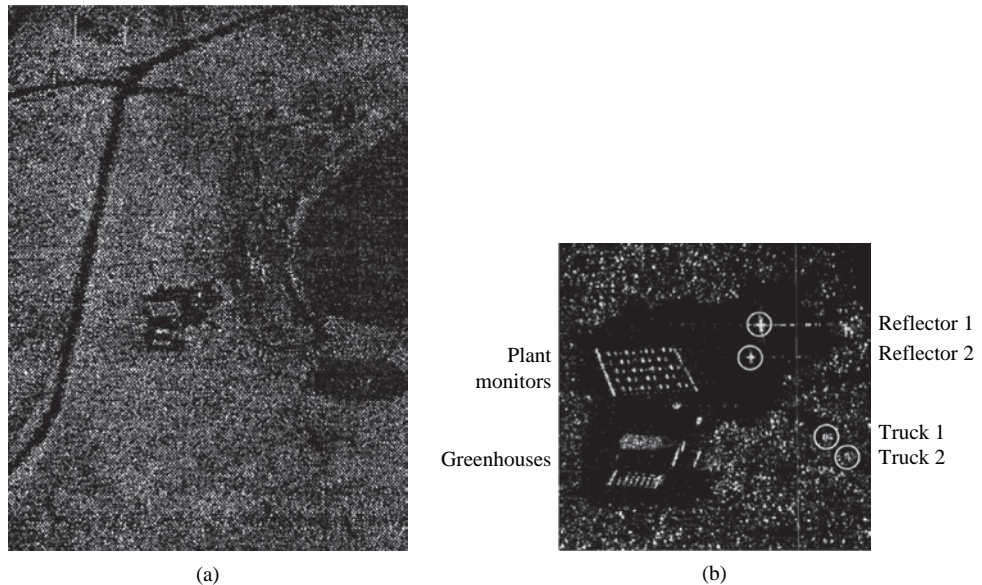
A common theme across these applications is the goal of surveying a large area; indeed, global coverage is sometimes desirable. The properties of an isolated 1 km by 1 km spot on the earth are rarely of interest. Consequently, spotlight operation, with its limited image extents and low area coverage rates, is quite unusual for remote sensing. On the other hand, moderate cross-range resolutions are required to segment and isolate different terrain types and facilitate parameter retrieval as a function of location. Therefore, ScanSAR collections, with their associated poor cross-range resolutions, have limited utility. Stripmap possesses a reasonable balance between cross-range resolution and ACR, making it the standard acquisition mode for remote sensing.

The Canadian Space Agency's RADARSAT-1 is a fine example of a space-based SAR for remote sensing applications [18,19]. The radar antenna can be steered electronically in elevation, providing a ScanSAR capability. However, the antenna is incapable of either electronic or mechanical steering in azimuth, and continuous slewing of the platform is impractical; therefore, the system does not possess a spotlight mode. Stripmap operation produces the finest cross-range resolution, and RADARSAT is frequently tasked in this mode. Example imagery appears in Figure 7-60 [20].

7.7.2 Penetration Applications

It is sometimes desirable to produce imagery of objects through an intervening medium. Applications include the detection and characterization of ground vehicles hidden from view underneath treetop canopies, detection and localization of buried landmines and subterranean pipes, and the estimation of ground elevation beneath glacial ice sheets. The intervening medium in these three cases is tree foliage, earth soil and rocks, and glacial ice, respectively.

FIGURE 7-61 ■
 Example P-3 FOPEN SAR example of (a) a stripmap image of a forested region showing roads and clearing with a greenhouse near the middle, and (b) expanded view of the greenhouse clearing showing reflector calibration targets and two target vehicles under tree canopy foliage. (From Sheen et al. [23]. With permission of the IEEE.)



Medium penetration encourages the use of low frequencies and long wavelengths. The impact on SAR is twofold:

- A antenna aperture size that produces a narrow beamwidth at high frequencies will have a wide beamwidth at low, penetrating frequencies, a trend suggested by (7.113).
- Wide integration angles are required to achieve fine cross-range resolutions at long wavelengths, as indicated by (7.2).

These two observations were intertwined in [1] when we found that the stripmap integration angle is limited to the antenna beamwidth. In short, low frequencies provide long wavelengths for penetration but also lead to wide antenna beamwidths, but these same wide beams yield reasonable stripmap cross-range resolutions.

The ERIM/NAWC P-3 Ultra-Wideband SAR is a representative example of a FOPEN imaging system [21,22]. Example imagery appears in Figure 7-61 [23]. Stripmap operation yields acceptable cross-range resolutions that are consistent with the down-range resolution provided by the operating bandwidth; further improvements to cross-range resolution through implementation of a spotlight mode are unnecessary. Indeed, the beamwidth of the antenna system is already so wide that scanning the beam, in either a ScanSAR or spotlight mode, would have a negligible impact on resolution and ACR. For these reasons, low frequency SAR system modes tend to be exclusively stripmap in nature.

7.8 | SUMMARY

Stripmap SAR employs a linear collection geometry and a fixed antenna orientation, typically to broadside, that is, normal to the platform velocity vector. This mode of operation achieves the finest possible cross-range resolution while surveying the passing terrain without gaps in coverage.

The PSR, the manifestation of a point target in the raw data, takes the form of a hyperbolic range and phase history. The family of simple DBS image formation algorithms may be formulated using the hyperbolic PSR; however, these techniques are inadequate for high-fidelity stripmap imaging. RMA is the preferred technique for stripmap collections, as it is computationally efficient and makes no approximations to the PSR. RMA achieves its success by matched filtering through multiplication in the two-dimensional frequency domain, employing a one-dimensional Stolt interpolation in that same domain to overcome the down-range variation in the form of the PSR, and using the FFT to go to and from frequency domains. Other techniques like RSA and various forms of RDA were presented here for historical reasons and to buttress the development of RMA.

Along-track sampling requirements and range-ambiguous returns impose upper and lower limits to stripmap PRFs. These limits are set by azimuth and elevation beamwidths, which in turn are determined by antenna size. There is a minimum antenna size for meeting both upper and lower PRF limits, though these limits may be relaxed by sacrificing cross-range resolution or swath depth. Airborne systems are rarely challenged by these constraints, whereas space-based SAR designs are often driven by antenna size and PRF considerations. On the other hand, increasing antenna size reduces area coverage rate and spoils cross-range resolution. In short, oversized antennas limit system utility.

Remote sensing systems frequently prefer continuous coverage of a swath without sacrificing cross-range resolution, making stripmap the preferred mode of operation. Often these systems, particularly those in earth orbit, are incapable of beam steering, making stripmap the only mode option. The low operating frequencies of FOPEN antennas give rise to wide beamwidths, diminishing the benefit of beam steering for spotlight collections, so that these systems perform just as well in a stripmap mode.

7.9 | FURTHER READING

Stripmap SAR is a relatively mature discipline, so many of the seminal papers in the field were published quite some time ago. Unfortunately, age makes these papers difficult to read and understand due to an evolution in SAR and digital signal processing terminology over the past three decades, references to and descriptions of arcane and dead-end approaches to image formation, and, particularly in older papers, a lack of appreciation for the benefit of operating in the wavenumber (spatial-frequency) domain. Nevertheless, [3,7,10,11,13] bear mentioning. Ironically, the spotlight SAR book by Carrara et al. [11,13] possesses one of the most accessible descriptions of RMA and CSA available. The relationship between RMA and RDA, including modified RDA, is available in a unifying framework in Bamler's paper [10]. The mature textbook by Curlander and McDonough [3] provides an excellent discourse on SAR systems engineering in general and space-based SAR operations in particular. The chapters on image formation, however, are difficult to follow. Finally, while the journal paper by Munson and Visentin [7] is dated and coverage of image formation is limited, the authors emphasize a signal processing point of view. Fundamental aspects of SAR data collection and processing are described in detail, and airborne and space-based operations are compared and contrasted in an informative manner.

7.10 | REFERENCES

- [1] Showman, G. A., "An Overview of Radar Imaging," Chapter 21 in Richards, M.A., Scheer, J.A., and Holm, W.A., Eds., *Principles of Modern Radar: Basic Principles*, SciTech Publishing, Raleigh, NC, 2010.
- [2] Sack, M., Ito, M.R., and Cumming, I.G., "Application of Efficient Linear FM Matched Filtering Algorithms to Synthetic Aperture Radar Processing," *IEEE Proceedings*, vol. 132, pt. F, no. 1, pp. 45–57, February 1985.
- [3] Curlander, J. and McDonough, R., "Other Imaging Algorithms," Chapter 10 in *Synthetic Aperture Radar: Systems and Signal Processing*, John Wiley & Sons Inc., New York, 1991.
- [4] Soumekh, M., "Generic Synthetic Aperture Radar," Chapter 4 in *Synthetic Aperture Radar Signal Processing*, John Wiley & Sons Inc., New York, 1999.
- [5] Stoker, J.J., "Unsteady Motion," Chapter 6 in *Water Waves: The Mathematical Theory with Applications*, Interscience Publishers Inc., New York, 1957.
- [6] Curlander, J. and McDonough, R., "Imaging and the Rectangular Algorithm," Chapter 4 in *Synthetic Aperture Radar: Systems and Signal Processing*, John Wiley & Sons Inc., New York, 1991.
- [7] Munson, D.C. and Visentin, R.L., "A Signal Processing View of Strip-Mapping Synthetic Aperture Radar," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 2131–2147, December 1989.
- [8] Thompson, A.A., Cheung, E.S.H., and Chang, C.Y., "Precision SAR Processing for RADARSAT," in *Proceedings of IGARSS'95: The International Geoscience and Remote Sensing Symposium*, vol. 3, pp. 2307–2309, 1995.
- [9] Cafforio, C., Prati, C., and Rocca, E., "SAR Data Focusing Using Seismic Migration Techniques," *IEEE Trans. AES*, vol. 27, no. 2, pp. 194–207, March 1991.
- [10] Bamler, R., "A Comparison of Range-Doppler and Wavenumber Domain SAR Focusing Algorithms," *IEEE Trans. GRS*, vol. 30, no. 4, pp. 706–713, July 1992.
- [11] Carrara, W.G., Goodman, R.S., and Majewski, R.M., "Range Migration Algorithm," Chapter 10 in *Spotlight Synthetic Aperture Radar*, Artech House, Boston, 1995.
- [12] Hughes, W., Gault, K., and Princz, G.J., "A Comparison of the Range-Doppler and Chirp Scaling Algorithms with Reference to RADARSAT," in *Proceedings of IGARSS'96: The International Geoscience and Remote Sensing Symposium*, vol. 2, pp. 1221–1223, 1996.
- [13] Carrara, W.G., Goodman, R.S., and Majewski, R.M., "Chirp Scaling Algorithm," Chapter 11 in *Spotlight Synthetic Aperture Radar*, Artech House, Boston, 1995.
- [14] Hawkins, D.W., "Synthetic Aperture Imaging Algorithms: With Application to Wide Bandwidth Sonar," Ph.D. thesis, University of Canterbury, Christchurch, New Zealand, October 1996. Available at http://ir.canterbury.ac.nz/bitstream/10092/1082/1/thesis_fulltext.pdf.
- [15] Freeman, A., Johnson, W.T.K., Huneycutt, B., Jordan, R., Hensley, S., Siqueira, P., and Curlander, J., "The 'Myth' of the Minimum SAR Antenna Area Constraint," *IEEE Trans. AES*, vol. 38, no. 1, pp. 320–324, January 2000.
- [16] Nguyen, L., Kapoor, R., and Sichina, J., "Detection Algorithms for Ultra-Wideband Foliage Penetrating Radar," in *Proceedings of SPIERadar Sensor Technology II*, vol. 3066, pp. 165–185, 1997.
- [17] NASA, *Spaceborne Synthetic Aperture Radar: Current Status and Future Directions*, Technical Memorandum 4679, April 1995.

- [18] Raney, R.K., Luscombe, A.P., Langham, E.J., and Ahmed, S., “RADARSAT,” *Proceedings of the IEEE*, vol. 79, no. 6, pp. 839–849, 1991.
- [19] Canadian Space Agency Website for RADARSAT-1. Available at <http://www.asc-csa.gc.ca/eng/satellites/radarsat1/>.
- [20] *EarthView Advanced Precision Processor (APP) for SAR*, Version 3.0 User’s Guide, Atlantis Scientific Inc., Nepean, Ontario, Canada, 2003.
- [21] Sheen, D.R., Strawitch, C.M., and Lewis, T.B., “UHF Wideband SAR Design and Preliminary Results,” in *Proceedings of IGARSS’94: The International Geoscience and Remote Sensing Symposium*, vol. 1, pp. 289–291, 1994.
- [22] Goodman, R., Tummala, S., and Carrara, W.G., “Issues in Ultra-Wideband, Widebeam SAR Image Formation,” *Record of the IEEE 1995 National Radar Conference*, pp. 479–485, 1995.
- [23] Sheen, D.R., VandenBerg, N.L., Shackman, S.J., Wiseman, D.L., Elenbogen, L.P., and Rawson, R.F., “P-3 Ultra-Wideband SAR: Description and Examples,” *IEEE AES Magazine*, vol. 11, pp. 25–30, 1996.

7.11 | PROBLEMS

You have been tasked to design a dual-frequency stripmap SAR system suitable for deployment on a small unmanned air vehicle and to implement and simulate image formation processing for this sensor. The two operating bands are L-band and X-band, with center frequencies of 1.0 GHz and 10 GHz, respectively. The specified down-range to scene center is 1 km, and the supported down-range and cross-range resolution is 1 m. A swath depth of 200 m is desired. Assume the platform flies at 100 m/s and the grazing angle at swath center is 30° . Use 3×10^8 m/s as the speed of light in your calculations.

1. [Basic calculations] What waveform bandwidth, antenna size (horizontal length), and along-track sampling interval are required for this system (same for both frequencies)? How much integration angle is required at L-band and X-band? Determine the approximate SAR baseline length using the arc length due to the integration angle.
2. [DBS-AD and RMC] Determine whether quadratic phase is significant by calculating the maximum down-range beyond which quadratic phase is a concern. If quadratic phase is significant, calculate the maximum phase deviation at the beginning (or, equivalently, the end) of the SAR baseline with respect to the middle. Find the amount of total slant-range change over a full-resolution stripmap collection by using the hyperbolic PSR to calculate slant range to a scatterer in the middle of the scene ($x = 0$, $r = r_0$) at the middle and ends of the SAR baseline. Compared with the down-range resolution, is migration of scatterers through range resolution cells a problem?
3. [RDA DOF] Calculate the depth of focus for the RDA technique employing the full PSR in the frequency domain. What is the impact on RDA in light of the desired swath depth?
4. [PRF limit] What is the lower PRF limit for full (finest) stripmap cross-range resolution? What is the corresponding swath limit? Is finding a good PRF a challenge for the desired swath depth?
5. [ACR and system utility] What is the achieved area coverage rate and system utility given the required swath depth? What is the potential system utility for finest cross-range resolution and processing of the full beam extent?

6. [MATLAB slant range and phase] Write MATLAB code to calculate the slant range to a scatterer at any desired down-range r and a cross-range x . Use $(x, r) = (0, r_0)$ for this exercise. Evaluate and plot the slant range for a set of along-track sample location separated by the along-track sampling interval and spanning the SAR baselines for the L-band and X-band systems. Multiply these slant-range vectors by $4\pi/\lambda_c$ to generate the phase history at their corresponding center frequencies. Plot these phase histories.
7. [MATLAB signal model] Extend the one-dimensional phase histories produced in problem 6 to two dimensions, where the second dimension is frequency. Sample discretely about the L-band and X-band center frequency over an extent that covers the required waveform bandwidth and by a uniform interval equal to 100 kHz. Plot these frequencies. Take the resulting two-dimensional (frequency and along-track position) phase histories and convert them to complex signal histories using $\exp(-j*\text{phase})$. Ignore any power variations due to range-to-the-fourth, antenna gain patterns, and so on, by using a magnitude of 1.0. Generate an image of the phase angle of each signal. (At this point you have a simple and generic SAR data generation environment capable of producing stepped-frequency records as a function of along-track position of the platform. You may vary center frequency, bandwidth, frequency sampling, SAR baseline length, along-track sampling, and range to scene center to perform parametric analyses. Another loop may be added to coherently sum the returns from multiple scatterers in the scene.)
8. [MATLAB pulse compression] Use IFFT over frequency to generate pulse-compressed measured data as a function of along-track location and time delay, $d(u, t)$. Create an image of the magnitude of the L-band and X-band data. Depict the phase by creating an image of the real part of the pulse-compressed data. (The image of the real data is easier to interpret if a linear phase progression over time is detrended. This is easily accomplished by first IFFTSHIFTing over frequency prior to the IFFT from frequency to time. Multiplying the data by a phase offset of -90° , or $-1j$, improves interpretability further.)
9. [DBS imaging] Generate the range-compressed data as in problem 8 omitting the IFFTSHIFT and the phase offset described toward the end of the problem. Perform DBS imaging as documented in Figure 7-5. Begin by Fourier transforming over the along-track position with an FFT and then place zero cross-range in the middle of the image by applying an FFTSHIFT over cross-range. Index pixels to their raw (k_u, ω) locations initially, then use the mapping in the text to locate pixels to cross-range/down-range (x, r) .
10. [DBS-AD imaging] Implement DBS with AD as documented in Figure 7-10 and compare the resulting imagery with DBS imagery for L-band and X-band.
11. [RDA PSF] Generate the frequency-domain (k_u, ω) form of the PSR for the swath center r_0 for the L-band and X-band cases. Use the RF support (vector) corresponding to the raw data generated in problem 7 and the spatial-frequency support (vector) created in problem 9. Produce phase images of these PSRs. Pulse-compress the PSRs as in problem 8 by inverse Fourier transforming (using the IFFT) from frequency to time and create images of the Doppler-domain time-delay profiles. (Recall this form of the PSR has been offset to put the closest point of approach at down-range r_0 at time delay equals zero, so the output should be FFTSHIFTed and the time-delay axis recalculated accordingly.) Finally, examine the PSR as a function of (u, t) by inverse

Fourier transforming (using the IFFT) from k_u to u . FFTSHIFT to place the closest point of approach in the middle. (Indexing the u axis in meters may be a challenge if the k_u sampling from problem 9 was used to generate the PSR and oversampling was used to go from u to k_u in problem 9.)

12. [RDA imaging] Perform RDA imaging by matched filtering with the full PSR (via multiplication in the frequency domain) as documented in Figure 7-44. Start with the raw, frequency-domain data generated in problem 7, as they should already have the same RF sampling as the PSR created in problem 11. Fourier transform the synthetic data from u to k_u (using the FFT) with the appropriate amount of zero padding, if any, and FFTSHIFT to register the data with the PSR from problem 11. (If the k_u domain is oversampled, there will be a linear phase modulation on the data over spatial frequency that will cause an undesirable cross-range offset in the final RDA image. You will have to estimate and compensate for this phase by multiplying the data by the conjugate of your estimate.) Complete the RDA processing using a two-dimensional IFFT and FFTSHIFTING in cross-range. (The point scatterer will, hopefully, be focused for both L-band and X-band, but perhaps not at the appropriate down-range/cross-range location. Accounting for image offsets due to FFT-induced linear phase modulations, especially when zero padding is involved, is notoriously challenging. Keep in mind that erroneous shifts in (x, r) are due to uncompensated linear phase in (k_u, ω) .)

Interferometric SAR and Coherent Exploitation

Mark A. Richards

Chapter Outline

8.1	Introduction	337
8.2	Digital Terrain Models	342
8.3	Estimating Elevation Profiles Using Radar Echo Phase	344
8.4	InSAR Operational Considerations	359
8.5	InSAR Processing Steps	362
8.6	Error Sources	375
8.7	Some Notable InSAR Systems	382
8.8	Other Coherent Exploitation Techniques	386
8.9	Summary	392
8.10	Further Reading	392
8.11	References	393
8.12	Problems	397

8.1 | INTRODUCTION

Interferometric synthetic aperture radar (InSAR, also abbreviated as IFSAR¹) is a family of techniques that analyze phase differences between two or more complex (amplitude and phase) high-resolution synthetic aperture radar (SAR) images to gain information beyond just the reflectivity of scatterers in the scene. This information may be scatterer elevation, elevation change, or reflectivity change. InSAR is a subset of the more general topic of coherent exploitation of complex SAR images, which includes such additional techniques as along-track interferometry (ATI) and ground or surface moving target indication (GMTI or SMTI).

InSAR is most often used to generate high-quality terrain elevation maps, called digital elevation maps or sometimes digital elevation models (DEMs). The fine spatial resolution of SAR imagery enables measurements of terrain elevation on a dense grid of sample points, while the use of phase-based measurements at microwave frequencies enables good elevation accuracies of ones to tens of meters. Furthermore, the use of active microwave radar as the sensor inherently provides an all-weather, day–night capability to

¹Both acronyms are in wide use, but InSAR is the more common.

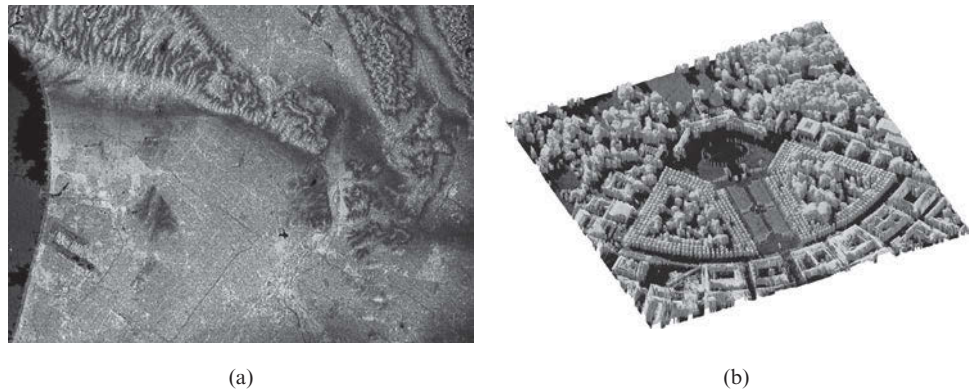


FIGURE 8-1 ■ Example InSAR images. (a) Spaceborne image of Los Angeles basin from SRTM data. Grayscale represents wrapped height. (Courtesy of NASA/JPL-Caltech.) (b) 3-D rendering of a high resolution airborne InSAR image of Karlsruhe Palace. (From Brenner and Roessing [1]. With permission.)

generate DEMs. Both spaceborne and airborne InSAR systems are in common use, and InSAR elevation maps can be generated from either stripmap or spotlight SAR imagery. Figure 8-1 illustrates two examples of 3-D scene renderings generated from InSAR DEMs. Figure 8-1a is a visualization of the Los Angeles area generated from space shuttle data. The image is 41 km wide by 29 km tall (25×19 miles) and has a resolution of 30 m. The flat Los Angeles basin is in the center and lower left of the image. The runways of Los Angeles International airport are visible, as is the Marina del Rey just to the north of the airport. The Santa Monica and Verdugo mountains run along the top of the image. The Pacific coastline is on the left. In this image, the grayscale is related to scatterer height rather than reflectivity.² Figure 8-1b shows a 3-D rendering of an airborne InSAR DEM of the Karlsruhe Palace in Germany. The image covers about 10 km^2 at a resolution finer than 0.1 m and is an excellent example of very fine-resolution urban InSAR [1].

In this chapter we introduce the concepts, techniques, and applications of InSAR with an emphasis on DEM generation. After a brief introduction to DEMs, the fundamental equations relating interferometric phase measurement variations to terrain elevation variations are derived from simple geometric considerations. The major algorithmic steps required to form an InSAR DEM are then discussed, and a number of existing air- and space-based InSAR systems are described.

In addition to DEM generation, the closely related coherent SAR exploitation techniques of terrain motion mapping and change detection are briefly introduced in this chapter. They provide high-quality measurements of changes in, respectively, the terrain profile over time or the reflectivity of the terrain. Another coherent exploitation technique, ATI, uses similar methods to detect terrain motion and moving targets in a scene. The basic concepts of ATI can be expanded to combine SAR imaging with space-time adaptive processing (STAP) methods for interference suppression, an idea developed further in Chapters 10 and 11 of this book.

²Although not available here, InSAR images typically rely on pseudo-color mappings to make the height detail more perceptible to the eye. Color versions of these and other selected images from this chapter are available online (<http://www.scitechpub.com/pomr2>).

8.1.1 Organization

This chapter begins with an introduction to digital elevation models, the desired output of the InSAR process, in Section 8.1.3. Section 8.3 discusses how elevation is measured in InSAR using pairs of range measurements. Analysis of the accuracy requirements establishes the need for using phase to measure range. The differences between one-pass and repeat-pass operation and between airborne and spaceborne InSARs are considered in Section 8.4.

Section 8.5 introduces the major processing steps needed to actually form DEMs using InSAR, with an emphasis on the problem of phase unwrapping in two dimensions. Brief discussions of image registration, absolute elevation estimation, and orthorectification and geocoding are included. Error sources in InSAR are discussed in Section 8.6, which also introduces the important concept of coherence. Selected notable airborne and spaceborne InSAR systems are described in Section 8.7. Although this chapter focuses strongly on 3-D InSAR elevation profile mapping, Section 8.8 provides a short introduction to the related methods of terrain motion mapping, change detection, and along-track interferometry. Finally, Sections 8.9 and 8.10 summarize the chapter and provide suggestions for further reading.

8.1.2 Key Points

- Coherent imaging radars, both airborne and spaceborne, can generate very high-quality DEMs covering large areas on dense sampling grids at day or night and in all weather.
- Terrain elevation profile variations can be measured using pairs of range measurements, but achieving sufficient elevation profile accuracy requires that range be measured to small fractions of a typical radar wavelength; hence, radar echo phase variations are used to measure small range variations.
- Formation of InSAR elevation maps requires formation of pairs of high-quality SAR images, followed by a complicated series of signal processing operations to develop the DEM such as subpixel image registration, two-dimensional phase unwrapping, and orthorectification to correct geometric distortions.
- A variety of sensor configurations and operational concepts can be used to acquire InSAR data, including single- and multiple-receiver radars, one- and repeat-pass collection protocols, and multiplatform formation flying protocols.
- The coherence of an InSAR image pair is the primary determinant of the elevation estimate quality. It is affected by a variety of statistical and systematic error sources.
- A variety of airborne and spaceborne InSAR systems have been placed in service, most since about 1995. New systems continue to be developed and refined.
- Using appropriate sensor configurations, the same basic technology can be used to map scatterer and terrain motion and reflectivity changes over time scales from fractions of a second to years.

8.1.3 Notation

Following are the principal notations used in this chapter:

- x^* complex conjugate of x
- $\langle \cdot \rangle_{2\pi}$ evaluation modulo 2π
- α_{IF} interferometric scale factor

β	sensor tilt angle, relative to horizontal
$\widehat{\delta h}$	estimated change in height
δR	difference in range to two sensors
$\tilde{\Delta}_x, \tilde{\Delta}_y$	x - and y -dimension gradients of the interferometric phase difference
ϕ	phase
ϕ_a, ϕ_b	phase shift at receivers a and b , respectively
ϕ_{ab}	interferometric phase difference
$\phi_G, \phi_R, \phi_\rho$	phase shift due to radar receiver transfer function, propagation range, and scatterer complex reflectivity, respectively
ϕ_{ab}^{FE}	flat-earth interferometric phase difference
$\tilde{\phi}, \tilde{\phi}_a, \tilde{\phi}_b, \tilde{\phi}_{ab}$	wrapped values of phases $\phi, \phi_a, \phi_b, \phi_{ab}$
$\tilde{\phi}_{ab}^{ref}$	wrapped reference interferometric phase difference
$\tilde{\phi}_{ab}^{dif}, \hat{\phi}_{ab}^{dif}$	wrapped, estimated differential interferometric phase differences
Φ	objective function for phase unwrapping
γ	coherence
$\gamma_{noise}, \gamma_{spatial}, \gamma_{temporal}$	coherence due to noise, baseline decorrelation, and temporal decorrelation, respectively
λ	wavelength
σ_h	standard deviation of h
ξ	peak-to-RMS ratio
ψ	general depression angle relative to horizontal
ψ_0	depression angle to foreshortened scatterer, relative to horizontal
ψ_1	depression angle to elevated scatterer, relative to horizontal
ψ_B	angular baseline (radians)
$\sigma_{\phi_{ab}}$	standard deviation of interferometric phase difference
$\sigma_{\hat{\phi}_{ab}}^2$	variance of estimated interferometric phase difference
$\sigma_{\hat{\gamma}}^2$	variance of estimated coherence magnitude
B	sensor baseline separation
B_t	temporal baseline
B_{tc}	decorrelation time of temporal baseline
B_\perp	sensor baseline projected perpendicular to the line of sight
$B_{c\perp}$	critical perpendicular baseline
d	driving function for Ghiglia-Romero phase unwrapping algorithm
D	two-dimensional discrete cosine transform of d
$D_{t\perp}$	tomographic aperture projected orthogonal to the line of sight
h	scatterer elevation
\hat{h}	estimated elevation
\hat{h}_{dif}	estimated differential elevation
H	platform height above ground level
I_{ab}	interferogram
p	single- versus dual-transmitter scale factor for interferometric phase difference
P	error power in phase unwrapping objective function
$\mathbf{P}_1, \mathbf{P}_0$	actual and foreshortened scatterer locations

R	range
s_{ab}^n	correlation function between n -th subimages of images a and b
v_r	radial velocity
w_x, w_y	weighting functions for phase unwrapping
x	cross-range (along-track) coordinate
y, y_1	ground range (cross-track) position
y_0	foreshortened ground range position
Δy	foreshortening or layover in ground range coordinate
z	elevation coordinate

8.1.4 Acronyms

Following are the definitions of the acronyms used in this chapter.

ATI	along-track interferometry
CCD	coherent change detection
CRLB	Cram�r-Rao lower bound
CTI	cross-track interferometry
DCT	discrete cosine transform
DEM	digital elevation map, digital elevation model
DFT	discrete Fourier transform
DHM	digital height model
DLR	German Aerospace Center
DSM	digital surface model
DTED	digital terrain elevation data
DTM	digital terrain model
ERIM	Environmental Research Institute of Michigan
ERS	European remote sensing
ESA	European space agency
E-SAR	European synthetic aperture radar
GMTI	ground moving target indication
GZW	Goldstein-Zebker-Werner
HRE	high-resolution elevation
HRTI	high-resolution terrain information
IFSAR	interferometric synthetic aperture radar
InSAR	interferometric synthetic aperture radar
InSARE	InSAR elevation
IPD	interferometric phase difference
JPL	Jet Propulsion Laboratory
LIDAR	light detection and ranging
LOS	line of sight
MCF	minimum cost flow
NASA	National Aeronautics and Space Administration
NGA	National Geospatial intelligence Agency
PDF	probability density function
PRI	pulse repetition interval
RF	radar frequency
RMS	root mean square
RTV	rapid terrain visualization
SAR	synthetic aperture radar

SIR-C	shuttle imaging radar C
SMTI	surface moving target indication
SRTM	shuttle radar topography mission
STAP	space-time adaptive processing
UTM	universal transverse Mercator
WGS	world geodetic system

8.2 | DIGITAL TERRAIN MODELS

A digital terrain model (DTM) is a digital representation of the elevation of a portion of the Earth's surface [2]. It typically is composed of elevation measurements for specified points, lines, and surface elements and may also include an interpolation scheme for estimating elevation between sample points and descriptive metadata such as the latitude and longitude of a cell, its estimated height error, the data source, and so forth. The term DEM, also sometimes called a digital height model (DHM), usually implies a simple array of regularly spaced elevation values referenced to a standard geographic coordinate system [3]. A DHM or DEM is a special case of a DTM. The term DEM also refers to a specific class of data products available from the U.S. Geological Survey. The data in a DTM or DEM are usually intended to represent the elevation of the bare or bald earth. In contrast, a digital surface model (DSM) is a representation of the top of the terrain rather than the bare earth. In a forested area, a DSM estimates the elevation of the forest canopy, while a DEM estimates the elevation of the forest floor.

Digital terrain models have an expanding variety of uses. The most obvious and important is topographic mapping, which in turn is useful for such diverse applications as three-dimensional visualization, terrain analysis for precision agriculture, line-of-sight (LOS) mapping for telecommunications tower siting and utilities routing, disaster analysis (e.g., flood mapping), and navigation. A less obvious example is the use of DTMs to enhance radar GMTI and STAP performance by incorporating knowledge of the terrain into the clutter statistics estimation procedures at the core of GMTI and STAP algorithms [4].

What degree of accuracy makes for a useful DEM? The quality of a DEM is determined by the spacing of the grid points (the denser, the better) and the accuracy of the individual elevation values. A commonly cited DEM standard is the Digital Terrain Elevation Data (DTED) specification developed by the U.S. National Geospatial Intelligence Agency (NGA) and its predecessors [5]. DTED DEM data are classified into six DTED levels numbered 0 through 5, though only levels 0 through 2 are standardized [6]. Table 8-1 shows the

TABLE 8-1 ■ Selected DTED Specifications

DTED Level	Post Spacing [‡]	Absolute Vertical Accuracy	Relative Vertical Accuracy
0	30.0 arc sec \approx 928 m	< 30 m	< 20 m
1	3.0 arc sec \approx 93 m	< 30 m	< 20 m
2	1.0 arc sec \approx 31 m	< 18 m	< 12–15 m
3 [†]	0.3333 arc sec \approx 10 m	< 10 [†] m	< 1–3 [†] m
4 [†]	0.1111 arc sec \approx 3 m	< 5 [†] m	< 0.8 [†] m
5 [†]	0.0370 arc sec \approx 1.1 m	< 5 [†] m	< 0.33 [†] m

[†]Accuracies for DTED levels 3–5 are proposed but not final and not included in MIL-PRF-89020B. Various sources report varying values for the proposed accuracy.

[‡]Post spacings in meters are based on a nominal earth radius of 6378.1 km.

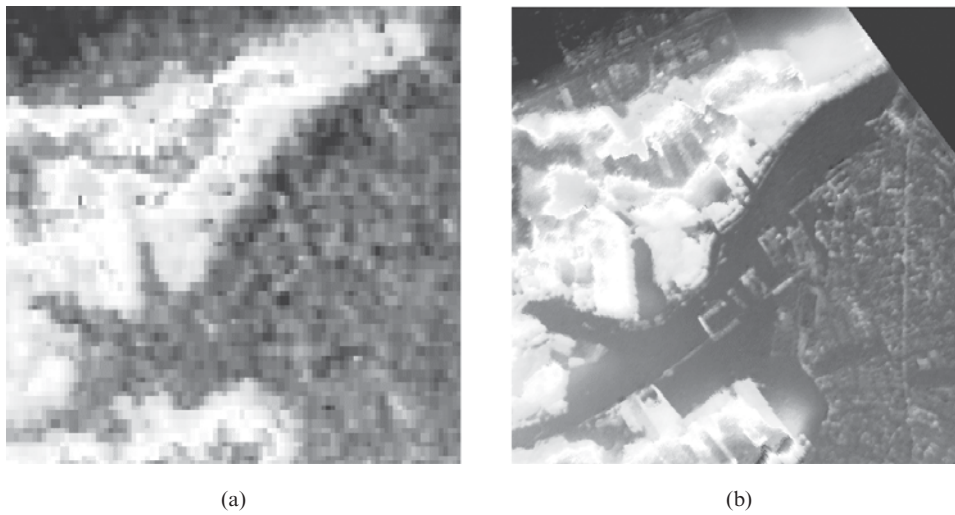


FIGURE 8-2 ■ Comparison of two DEMs of a 2×2 km area of Eichenau (close to Munich, Germany) derived from different DTED data levels. (a) Derived from DTED level 2 SRTM data. (b) Derived from DTED level 3 E-SAR data. (Courtesy of German Aerospace Center (DLR).)

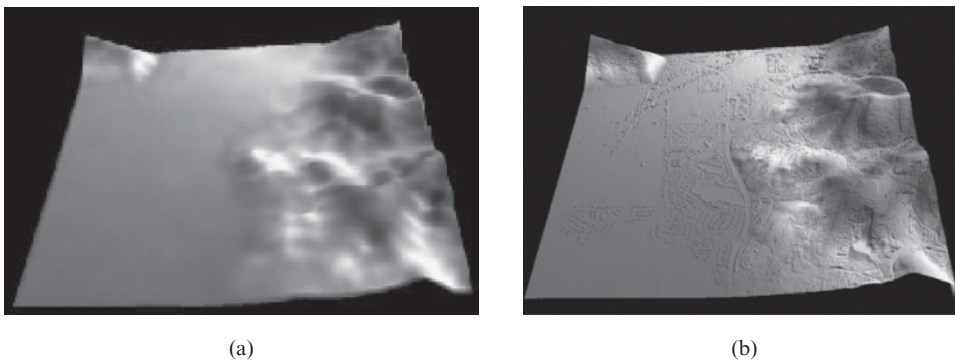


FIGURE 8-3 ■ Comparison of two DEMs of the Park City, Utah area. (a) DTED level 2 DEM from the U.S. Geological Survey. (b) DTED level 4 DEM from the Sandia RTV InSAR. (Courtesy of Sandia National Laboratories.)

increasing level of detail associated with increasing DTED levels [6-8]. The Shuttle Radar Topography Mission (SRTM) conducted in 2000 collected data from low Earth orbit intended to support mapping of 80% of the earth's surface at DTED level 2 [9,10]. The U.S. Army's Rapid Terrain Visualization (RTV) demonstration developed an airborne system for near real-time generation of DTED Level 3 and 4 products over localized areas [7]. To date, DTED level 5 data typically requires use of an airborne LIDAR system. Figure 8-2 compares a small portion of a DEM of the same area rendered at DTED level 2 derived from SRTM data (Figure 8-2a) with a DTED level 3 rendering, derived from E-SAR data (Figure 8-2b).ⁱ Figure 8-3 is another example comparing a U.S. Geological Survey DTED level 2 DEM of the Park City, Utah area to a DTED level 4 DEM based on data collected by the RTV InSAR system.

ⁱE-SAR is the "experimental SAR" developed by the German Aerospace Center (DLR). See http://www.op.dlr.de/ne-hf/projects/ESAR/esar_englisch.html.

TABLE 8-2 ■ Selected HRE Specifications

HRE Level	Product Name	Post Spacing	Absolute Vertical Accuracy LE90 [†] (goal)	Relative Vertical Accuracy
1	HREGP	0.4 arc-sec (≈ 12.4 m)	12.4 m	6.2 m
2	HRE80	8 m	8 m	4 m
3	HRE40	4 m	4 m	2 m
4	HRE20	2 m	2 m	1 m
5	HRE10	1 m	1 m	0.5 m
6	HRE05	0.5 m	0.5 m	0.25 m
7	HRE02	0.25 m	0.25 m	0.12 m
8	HRE01	0.125 m	0.12 m	0.06 m

[†]LE90 = linear error at 90% probability level. See [11].

Development of an updated version of the DTED specification, called high-resolution terrain information (HRTI), was initiated by NGA as standard MIL-PRF-89048. However, more recently NGA has superseded that effort with development of the high-resolution elevation (HRE) DEM standard [11], which provides eight levels and is suitable for newer, finer resolution sensors. Table 8-2 lists some of the HRE specifications. NGA has stated that it will continue to support DTED levels 0, 1, and 2 but not levels 3–5 or the HRTI draft specification. DTED levels 0–2 and HRE levels 1–8 collectively cover relative (point-to-point) vertical accuracies ranging from approximately 20 m to 0.06 m (about 2.4 inches).

8.3 | ESTIMATING ELEVATION PROFILES USING RADAR ECHO PHASE

SAR systems form fine resolution 2-D images using various approximations to matched filtering of the point target response of terrain scatterers, as discussed in the previous two chapters. In InSAR, the third dimension of elevation is generally obtained not by matched filtering, but by phase interferometry techniques. Some newer, more advanced systems are beginning to explore matched filtering in elevation, as described in Section 8.3.8.

8.3.1 The Effect of Scatterer Elevation on 2-D SAR Images

A general stereo imaging geometry is shown in Figure 8-4. For simplicity, we consider side-looking radar only (no squint). A sensor platform such as an aircraft or spacecraft travels in the x -dimension (normal to the plane of the figure). A coherent pair of imaging sensors, denoted a and b , are separated by a *spatial baseline* B orthogonal to the velocity vector (thus in the y - z plane) at an angle β with respect to horizontal. The angle β is often approximately either 0° (horizontal displacement of sensors) or 90° (vertical displacement). The sensors are located at an arbitrary position x_0 in the cross-range (also called along-track) coordinate x , which is not shown in the figure. The center of the baseline is located at an elevation $z = H$ relative to the x - y ground plane and ground range position $y = 0$.³

³We assume a flat earth (flat ground plane) for simplicity. In practice, spaceborne InSAR systems must account for Earth curvature.

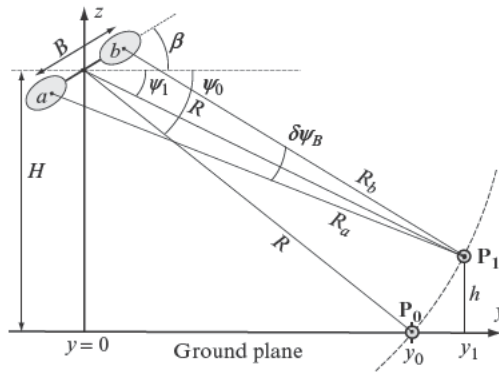


FIGURE 8-4 ■
Stereo imaging
geometry.

2-D SAR images are formed in a specified plane such as the slant range plane or, more likely, in the ground plane. Consider a pixel located at ground range position y_0 in a 2-D SAR image formed in the ground plane using an appropriate stripmap or spotlight SAR algorithm.⁴ We assume we know the value of y_0 to within the image ground range resolution ΔR , the complex reflectivity $\rho = Ae^{j\phi}$ of that pixel, and the sensor pair altitude H , baseline length B , and baseline tilt β . From this data we can compute the slant range and depression angle from the baseline center to the pixel using

$$R = \sqrt{H^2 + y_0^2}$$

$$\psi_0 = \tan^{-1} \left(\frac{H}{y_0} \right) = \sin^{-1} \left(\frac{H}{R} \right) = \cos^{-1} \left(\frac{y_0}{R} \right) \quad (8.1)$$

However, *any* point in the y - z plane that is located on the isorange contour passing through y_0 will also be imaged at y_0 in the 2-D SAR image. The physical scatterer imaged at y_0 might in fact be \mathbf{P}_0 , located at (y,z) coordinates $(y_0, 0)$, but it might also be \mathbf{P}_1 , located at (y,z) coordinates (y_1, h) .⁵ The depression angles from the baseline center to these actual scatterer locations are ψ_0 and ψ_1 .

How can we distinguish these two situations? If it is possible to directly measure the actual depression angle ψ_1 to the scatterer that resulted in the image at $(y_0, 0)$, then it is possible to determine the true ground range of the scatterer and thus the true elevation h , using

$$H - h = R \sin \psi_1 \quad \Rightarrow \quad h = H - R \sin \psi_1 \quad (8.2)$$

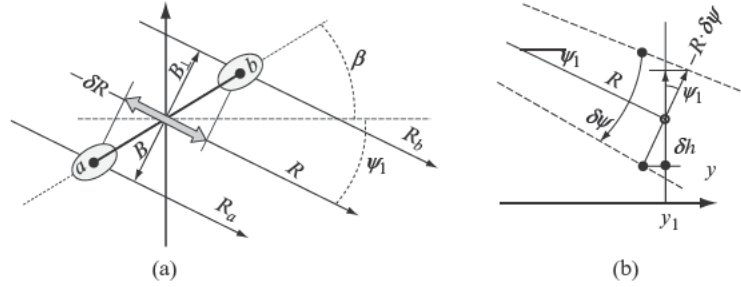
The problem is now one of how to measure ψ_1 .

8.3.2 Measuring Depression Angle with Pairs of Range Measurements

Referring again to Figure 8-4, assume that the scatterer responsible for the image pixel at location \mathbf{P}_0 is in fact \mathbf{P}_1 . Instead of deducing R from y_0 and H , we now measure the two

⁴An introduction to radar imaging and image formation algorithms is given in Chapter 21 of [12].

⁵The pixel at y_0 could also have contributions from both \mathbf{P}_0 and \mathbf{P}_1 , if the terrain elevation at y_0 and y_1 is such that both project to y_0 . We will assume that each image pixel is dominated by the reflectivity of one unique scatterer position.

FIGURE 8-5 ■(a) Geometry for approximating δR .(b) Geometry for approximating $d(\delta R)/d\psi$.

slant ranges R_a and R_b from the sensors a and b to \mathbf{P}_1 . Under the reasonable assumption that $R \gg B$, the lines of sight from a and b to \mathbf{P}_1 are very nearly parallel with the line of sight from the center of the baseline to \mathbf{P}_1 . This assumption greatly simplifies the derivations to follow. The approximate formulas developed here based on this assumption are of sufficient accuracy for most radar work. However, more precise but more complicated trigonometric relations can be used to solve for R and ψ if needed.⁶

As shown in Figure 8-5a, under this assumption R_a and R_b are well approximated as

$$\begin{aligned} R_a &\approx R + \frac{B}{2} \cos(\psi_1 + \beta) \\ R_b &\approx R - \frac{B}{2} \cos(\psi_1 + \beta) \end{aligned} \quad (8.3)$$

It follows from (8.3) that $R \approx (R_a + R_b)/2$. It will prove useful to work with the range difference $\delta R \equiv R_a - R_b$ instead of R_a and R_b directly. From equation (8.3), δR is

$$\delta R \equiv R_a - R_b \approx B \cos(\psi_1 + \beta) \quad (8.4)$$

Equation (8.4) can be solved directly for the depression angle

$$\psi_1 = \cos^{-1} \left(\frac{\delta R}{B} \right) - \beta \quad (8.5)$$

This result can then be used to solve for h in equation (8.2).

The accuracy needed in estimating δR to meet a given specification on elevation accuracy, for example DTED level 2 or HRE level 1, can be found by estimating the differential $d(\delta R)/dh$. It is convenient to factor the differential into two components:

$$\frac{d(\delta R)}{dh} = \frac{d(\delta R)}{d\psi} \cdot \frac{d\psi}{dh} \quad (8.6)$$

From equation (8.4), the rate of change in δR with respect to depression angle is

$$\frac{d(\delta R)}{d\psi} = -B \sin(\psi + \beta) \quad (8.7)$$

⁶For instance, since all three sides of the triangle formed by R_a , R_b , and B are known, R can be found as the median (bisector of B) using standard trigonometric formulas.

Referring to Figure 8-5b and still assuming $R \gg B$, it can be seen that an increase in elevation of δh produces a decrease in angle of $\delta\psi$ that satisfies $R \cdot \delta\psi \cdot \cos\psi \approx -\delta h$, so that $d\psi/dh \approx -1/R \cos\psi$. Using these results in (8.6) gives

$$\frac{d(\delta R)}{dh} \approx \frac{B \sin(\psi + \beta)}{R \cos\psi} \quad (8.8)$$

The sensitivity δh of the estimated elevation to changes in the estimated differential range $\delta(\delta R)$ is then, for a fixed range R ,

$$\delta(\delta R) \approx \frac{B \sin(\psi + \beta)}{R \cos\psi} \delta h \Rightarrow \delta h \approx \left[\frac{R \cos\psi}{B \sin(\psi + \beta)} \right] \delta(\delta R) \quad (8.9)$$

The first form of equation (8.9) shows that a change in elevation of δh is multiplied by a factor on the order of (B/R) in its effect on the range difference δR .

Applied in reverse in the second form, equation (8.9) can be used to estimate the error in measuring h due to an error in measuring δR at a given range R . In this form, it shows that achieving good elevation accuracy (small δh) from significant stand-off ranges R requires a large baseline B , great precision in measuring δR (small $\delta(\delta R)$), or both. Optical stereo imaging systems, with their very fine resolution, can achieve good results with a stereo camera pair on a small baseline in one-pass operation. Conventional SAR-based stereo imaging systems must generally use repeat-pass operation with significant separation between the two tracks to obtain look angle differences from the two tracks to the terrain area of interest on the order of 10° to 20° [13]. However, for interferometric SAR such large baselines are not practical.

For spaceborne InSAR systems, the baseline is typically on the order of 100–200 m, though it can be as large as 1 or 2 km, while the altitude ranges from approximately 240 km (for the space shuttle) to 800 km (for low Earth orbit satellites), giving (R/B) on the order of 2500 to 7000. For airborne systems, the standoff range is usually on the order of 10 to 20 kilometers, but the baseline is typically only a few meters to a foot or less, so that again (R/B) is on the order of several thousand. Because of this large multiplier, it is necessary to have very small values of $\delta(\delta R)$ if elevation errors are to be acceptable. Thus, we need to be able to measure very small differences in range from the scatterer to the two receive apertures.

As an example, consider the DTED level 2 relative vertical accuracy requirement of 12 m. Assume that $\beta = \psi = 45^\circ$. The SRTM mission operated at an altitude of about 240 km. Ignoring earth curvature for simplicity, the slant range to the surface will be $R \approx 340$ km. The SRTM baseline was $B = 60$ m, so to meet the 12 m accuracy requirement, the range difference needs to be accurate to within 2.2 mm! Even with subpixel range tracking to 1/20 of a pixel, this is much finer than can be supported by SAR range resolutions. For instance, 1/20 of the SRTM range resolution of 15 m is 0.75 m, bigger than the desired 2.2 mm by a factor of 334.

8.3.3 Estimating Elevation Variations from Phase Measurements

The need for differential range measurements accurate to millimeters to achieve required elevation accuracies leads to the use of phase instead of time delay in radar interferometry, and is the reason that InSAR is a *coherent* SAR exploitation technology. Assume a coherent radar using appropriate waveforms and matched filtering. The phase of a SAR image pixel

formed at nominal range R can be expressed as

$$\phi = \phi_G + \phi_\rho + \phi_R = \phi_G + \phi_\rho - 4\pi \left(\frac{R}{\lambda} \right) \quad (8.10)$$

where R should be considered the slant range to the pixel at the center of the synthetic aperture, ϕ_G is the phase shift due to the complex gain of the radar receiver and the image formation processor, ϕ_ρ is the net phase of the composite reflectivity of the scatterers contributing to the pixel, and $\phi_R = -4\pi R/\lambda$ is the deterministic phase shift due to the two-way path length to the pixel center [13,14]. For the sidelooking configuration assumed here, R is the slant range in the y - z plane as shown in Figure 8-4.

Let ϕ_a and ϕ_b be the pixel phases measured at the same ground range position (x_0, y_0) in the SAR images formed by sensors a and b in Figure 8-4. We assume the ϕ_G and ϕ_ρ terms are nearly equal in both images.⁷ It follows from equations (8.4) and (8.10) that

$$\phi_{ab} \equiv \phi_a - \phi_b = -4\pi \left(\frac{\delta R}{\lambda} \right) \approx -\frac{4\pi B}{\lambda} \cos(\psi + \beta) \quad (8.11)$$

Equation (8.11) applies when the radars at locations a and b are completely separate; each has its own transmitter and receiver, so that the two-way path lengths for the range measurements differ by δR . An important variation exists in the form of systems that use only one transmitter, say at location b , with two receivers at a and b . That is, the radar at b is an active system, whereas the radar at a is a passive radar receiver. In this event, the difference in the two-way path length observed at a and b is $\delta R/2$. Both cases can be combined into a single expression by replacing equation (8.11) with

$$\phi_{ab} = -2\pi p \left(\frac{\delta R}{\lambda} \right) \approx -\frac{2\pi p B}{\lambda} \cos(\psi + \beta) \quad (8.12)$$

where $p = 2$ when each sensor has its own transmitter, and $p = 1$ when one sensor serves as the transmitter for both. Operational scenarios that give rise to these cases are discussed in Section 8.4.1

The depression angle can now be estimated as

$$\psi = \cos^{-1} \left(-\frac{\lambda}{2\pi p B} \phi_{ab} \right) - \beta \quad (8.13)$$

The quantity ϕ_{ab} , called the *interferometric phase difference* (IPD), is a very sensitive indicator of range differences and thus of depression angle and elevation. The accuracy of IPD measurements is treated in Section 8.6, but after filtering for noise reduction the precision (standard deviation) is typically a few tens of degrees or better [15–17]. Therefore, ϕ_{ab} can be used to measure differential range with precisions of fractions of a radio frequency (RF) wavelength and thus enable much better elevation estimation accuracy. For example, a precision of 20° in ϕ_{ab} corresponds to a precision in measuring δR of $\lambda/36$, which is only about 8.3 mm at L band (1 GHz).

Once the elevation \hat{h} of the scatterer imaged at ground range y_0 has been estimated, the actual ground range y_1 of the scatterer can also be determined from the geometry in

⁷This is a good assumption for ϕ_G but might seem questionable for ϕ_ρ , particularly in repeat-pass InSAR (see Section 8.4.1). Variation of ϕ_ρ is one source of coherence loss, discussed in Section 8.6.1.

Figure 8-4. Because the angular change $\psi_1 - \psi_0$ is small, it is common to approximate the isorange arc connecting \mathbf{P}_0 and \mathbf{P}_1 as a straight line. Consideration of the triangle formed by \mathbf{P}_0 and \mathbf{P}_1 and the point at zero elevation and range y_1 then shows that

$$\Delta y \equiv y_1 - y_0 \approx h \tan \psi_1 \quad (8.14)$$

The shift Δy in the apparent ground range of the scatterer to y_0 from its actual ground range y_1 is called *foreshortening* or *layover*. The distinction is discussed in Section 8.3.7. Correction of foreshortening to form a new image with the scatterers located at the correct ground range is called *orthorectification* and is discussed in Section 8.5.8.

We now have our first method for estimating the elevation of the scatterer that accounts for the complex reflectivity of the SAR images at a ground plane pixel location (x, y) , assuming sidelooking operation. The observable data are the ground plane coordinates (x, y) and the complex reflectivities at each pixel. The algorithm is defined by the following procedure.

InSAR DEM Algorithm 1

1. For each ground plane image coordinate (x, y_0) , compute the range R to a scatterer of elevation zero at that location using equation (8.1).
2. Measure the IPD at that pixel from the two SAR images,
 $\phi_{ab}(x, y) = \phi_a(x, y_0) - \phi_b(x, y_0)$.
3. Use equation (8.13) to estimate the depression angle ψ_1 to the scatterer imaged at (x, y_0) .
4. Apply equation (8.2) to estimate the elevation \hat{h} of the scatterer.
5. Apply equation (8.14) to estimate the corrected ground range coordinate y_1 of the scatterer.

While this process seems straightforward, it does not estimate the actual elevation in practice because phase-based differential range measurements are highly ambiguous. Discussion of this *phase wrapping* problem and its solution is deferred to Section 8.3.4; we assume for now that an unambiguous estimate of \hat{h} is obtained by the InSAR method just described.

An alternative InSAR method is derived by considering the derivative of ϕ_{ab} with respect to δR in equation (8.12), leading to a relationship between a change in differential range and the corresponding change in IPD:

$$\delta \phi_{ab} = -\frac{2\pi p}{\lambda} \delta(\delta R) \quad (8.15)$$

Combining equations (8.9) and (8.15) gives a direct relationship between a change in scatterer elevation and the corresponding change in IPD:

$$\delta h \approx -\left[\frac{\lambda R \cos \psi}{2\pi p B \sin(\psi + \beta)} \right] \delta \phi_{ab} \equiv \alpha_{IF} \delta \phi_{ab} \quad (8.16)$$

where the *interferometric scale factor* α_{IF} has been defined. Note also that the length of the spatial baseline projected orthogonal to the LOS, denoted as B_\perp in Figure 8-5a, is $B \sin(\psi + \beta)$, so that an alternate expression for the interferometric scale factor is

$$\alpha_{IF} = -\frac{\lambda R \cos \psi}{2\pi p B_\perp} \quad (8.17)$$

Another alternate form is obtained in terms of the angular baseline $\delta\psi_B$ (see Figure 8-4), which satisfies $B_{\perp} \approx R \cdot \delta\psi_B$ so that

$$\alpha_{IF} = -\frac{\lambda \cos \psi}{2\pi p \cdot \delta\psi_B} \quad (8.18)$$

Typical values of $\delta\psi_B$ are quite small. For the earlier SRTM example, $\delta\psi_B \approx 177 \mu\text{rad}$ or about 0.01° . For a typical airborne InSAR case with $B_{\perp} = 1 \text{ m}$ and $R = 7 \text{ km}$, $\delta\psi_B \approx 143 \mu\text{rad}$ or about 0.008° .

Equation (8.16) states that a change in the elevation of the scatterer relative to the zero-elevation reference plane can be estimated by multiplying the measured change in the interferometric phase difference relative to the zero-elevation IPD by the interferometric scale factor.

To take advantage of this relationship, we first need the IPD map that results from a zero-elevation scene. Recasting equation (8.12) in terms of elevation and ground range, the IPD for a scatterer at zero elevation and ground range y becomes

$$\begin{aligned} \phi_{ab}(x, y) &\approx -\frac{2\pi p B}{\lambda} \cos[\psi(y) + \beta] \\ &= -\frac{2\pi p B}{\lambda} \cos\left[\cos^{-1}\left(\frac{y}{\sqrt{H^2 + y^2}}\right) + \beta\right] \quad (h(x, y) = 0) \\ &= -\frac{2\pi p B}{\lambda} \cos\left[\cos^{-1}\left(\frac{1}{\sqrt{1 + (H/y)^2}}\right) + \beta\right] \equiv \phi_{ab}^{FE}(x, y) \end{aligned} \quad (8.19)$$

This ground plane IPD is called the *flat earth IPD*, here denoted as ϕ_{ab}^{FE} . If we consider $\delta\phi_{ab}$ as the difference between the actual IPD and the flat earth IPD, then δh will be the difference between zero elevation and the actual elevation of the scatterer (i.e., h). The actual elevation of the scatterer producing the pixel at location (x, y) in the SAR images can therefore be estimated as

$$\hat{h}(x, y) \approx \alpha_{IF} [\phi_{ab}(x, y) - \phi_{ab}^{FE}(x, y)] \quad (8.20)$$

In using this equation, it is important to realize that the interferometric scale factor is a function of ψ and thus of ground range y . The value of α_{IF} must be updated for each ground range bin.

We now have a second InSAR DEM algorithm, defined by the following procedure.

InSAR DEM Algorithm 2

1. Using the two SAR images, measure the IPD $\phi_{ab}(x, y)$ at each ground plane pixel location (x, y) .
2. Compute $\phi_{ab}^{FE}(x, y)$ using equation (8.19).
3. For each ground plane image coordinate (x, y) , compute the range R to a scatterer of elevation zero at that location using equation (8.1).
4. Estimate the depression angle ψ of the scatterer imaged at each (x, y) ground coordinate using equation (8.13).

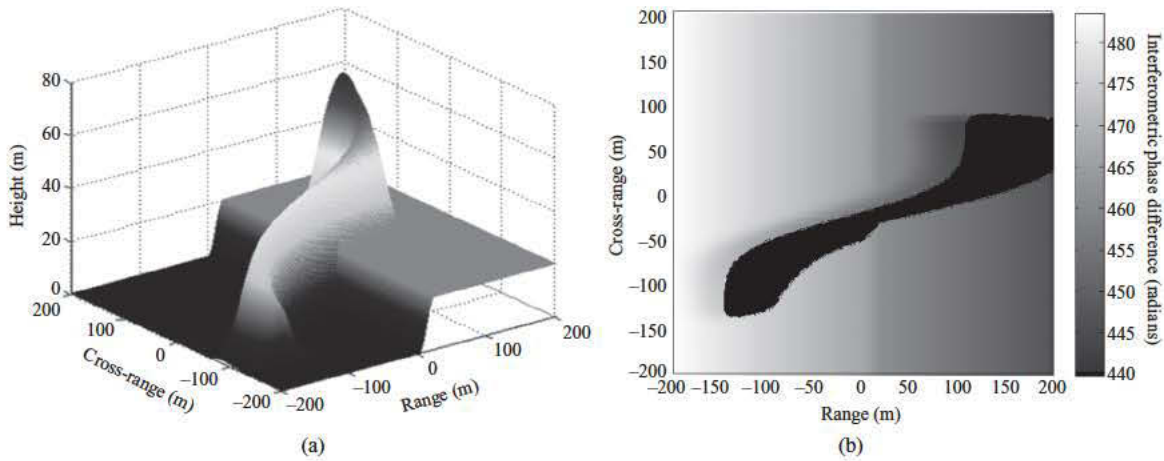


FIGURE 8-6 ■ (a) Notional elevation model. (b) Interferometric phase difference. See text for radar position and parameter details.

5. Compute α_{IF} using one of equations (8.16) through (8.18), as convenient.
6. Use ϕ_{ab} , ϕ_{ab}^{FE} , and α_{IF} in equation (8.20) to estimate the elevation profile $\hat{h}(x, y)$ of the scene.

Figure 8-6a shows a simple notional elevation model for demonstration purposes. The scene is 400×400 m in area with a peak elevation of 73.6 m. It includes a steep elevation plateau at a range of 0 meters (the center of the range interval) and an S-shaped rising “hill” running throughout the scene. The radar is located at a ground range of -5000 m and an altitude of 2000 m, making the depression angle and range from the center of the baseline to the center of the scene 21.8° and 5385 m, respectively. The two InSAR apertures are stacked vertically ($\beta = 90^\circ$) with a baseline of 3 m. The radar frequency is 10 GHz, yielding a wavelength of 3 cm, and the simulation assumes independent transmitters at each aperture (i.e., $p = 2$). The resulting interferometric scale factor at the center of the scene is $\alpha_{IF} \approx 4.29$.

Figure 8-6b shows the resulting IPD. The black area is the portion of the terrain that is shadowed by the hill. While the individual absolute phase shifts at apertures a and b are over 2 million radians, their difference varies from only about 443 to 483 radians over the scene. Equation (8.16) also predicts that the IPD will decrease as the elevation increases, an effect clearly visible at the plateau edge and on the hill surface. Note also that for constant-elevation regions, the IPD decreases as range increases as predicted by the flat earth IPD of equation (8.19). When the topography is estimated from this IPD using algorithm 1, the maximum absolute error in the unshadowed region is 0.42 m. When estimated using algorithm 2, it is 0.43 m. However, this is highly unrealistic, since we have not yet accounted for shadowing, phase wrapping, and noise. We will return to this example under more realistic conditions in Section 8.5.3.

8.3.4 Wrapped Phase

In any realistic scenario, the range R is many multiples of the wavelength λ , so the phase shift at each receiver is many multiples of 2π . For example, in the earlier Space Shuttle

example, the 340 km range is about 1,133,333 wavelengths at an RF of 1 GHz ($\lambda = 0.3$ m), so the phase shift is about $4,533,333\pi$ radians. An airborne X band (10 GHz, $\lambda = 3$ cm) radar at a range of 20 km experiences a phase shift of about $2,666,667\pi$ radians.

The phase ϕ_a of a complex signal sample $Ae^{j\phi_a} = I_a + jQ_a$, whether a time-domain sample or an image pixel, is measured in practice using the four-quadrant arctangent function $\text{atan}_4(Q_a/I_a)$ that returns not ϕ_a but its principal value $\tilde{\phi}_a$, which satisfies

$$\tilde{\phi}_a = \langle \phi_a \rangle_{2\pi} = \phi_a - k \cdot 2\pi \quad (8.21)$$

The notation $\langle \cdot \rangle_{2\pi}$ means modulo 2π . Here k is an unknown integer such that $\tilde{\phi}_a$ is in the range $[-\pi, \pi)$. $\tilde{\phi}_a$ is called the wrapped phase.

Given a wrapped phase measurement $\tilde{\phi}_a$, can we recover the original phase value ϕ_a ? In the absence of additional information, the answer for a single phase value is no. However, given a one-dimensional phase function $\phi_a(t)$, and assuming that $\phi_a(t)$ has finite bandwidth and is sampled at the Nyquist rate or above, it is possible to unwrap the sampled, wrapped measurements $\tilde{\phi}_a[n]$ to recover $\phi_a[n]$ [18]. Although there are practical difficulties, it is often possible to do so for the 2-D phase functions of InSAR, as will be discussed in Section 8.5.3. Thus, if we can compute the wrapped IPD $\tilde{\phi}_{ab}$ from the wrapped measurements $\tilde{\phi}_a$ and $\tilde{\phi}_b$, it should be possible to recover the unwrapped IPD ϕ_{ab} , which can then be used to determine the elevation profile.

One way to obtain $\tilde{\phi}_{ab}$ is to compute it as

$$\tilde{\phi}_{ab} = \langle \tilde{\phi}_a - \tilde{\phi}_b \rangle_{2\pi} \quad (8.22)$$

Note that $\tilde{\phi}_a - \tilde{\phi}_b$ can range from -2π to 2π , so equation (8.22) performs an additional wrapping operation on that result to wrap it into the range $[-\pi, \pi)$. To see that this produces the desired result, consider

$$\begin{aligned} \tilde{\phi}_a - \tilde{\phi}_b &= (\phi_a - 2\pi k_a) - (\phi_b - 2\pi k_b) \\ &= \phi_a - \phi_b + 2\pi(k_a - k_b) \\ &= \phi_a - \phi_b + 2\pi k \end{aligned} \quad (8.23)$$

for some k . Clearly $\langle \phi + 2\pi k \rangle_{2\pi} = \langle \phi \rangle_{2\pi} = \tilde{\phi}$ for any phase ϕ . Therefore

$$\langle \tilde{\phi}_a - \tilde{\phi}_b \rangle_{2\pi} = \langle \phi_a - \phi_b \rangle_{2\pi} = \tilde{\phi}_{ab} \quad (8.24)$$

Thus, the wrapped IPD can be computed by wrapping the difference between the wrapped phases at the individual apertures.

Another way to compute the wrapped IPD is to form the interferogram. Represent the two individual SAR images as $a(x, y) = A_a(x, y)e^{j\tilde{\phi}_a(x, y)}$ and $b(x, y) = A_b(x, y)e^{j\tilde{\phi}_b(x, y)}$. The interferogram is defined as

$$\begin{aligned} I_{ab}(x, y) &\equiv a(x, y)b^*(x, y) \\ &= A_a A_b \exp[j(\tilde{\phi}_a - \tilde{\phi}_b)] \end{aligned} \quad (8.25)$$

where b^* represents the complex conjugate of b . Consequently,

$$\begin{aligned} \arg\{I_{ab}(x, y)\} &= \text{atan}_4(\tilde{\phi}_a - \tilde{\phi}_b) \\ &= \tilde{\phi}_{ab} \end{aligned} \quad (8.26)$$

Note that the phase of the interferogram does not have to be explicitly wrapped again, as suggested by equation (8.22), because the only means we have to measure that phase

is to use the $\text{atan}_4(\cdot)$ function again, which returns a principal value and thus implicitly performs the additional wrapping step.

While the wrapped IPD of equation (8.26) is the primary function of interest, the amplitude weighting $A_a A_b$ of the full interferogram $I_{ab}(x, y)$ provides a measure of the signal-to-noise (SNR) of the data and therefore of the quality of the IPD estimate. This can be used to advantage in some phase unwrapping algorithms.

The need to rely on wrapped IPD rather than the true IPD to estimate elevation has implications that depend on the InSAR method being used. Consider the first InSAR method discussed earlier. There are now multiple possible solutions for the depression angle of equation (8.13), depending on the value of k :

$$\psi_k = \cos^{-1} \left(-\frac{\lambda}{2\pi p B} (\tilde{\phi}_{ab} + k \cdot 2\pi) \right) - \beta \quad (8.27)$$

For example, ψ_0 and ψ_1 in Figure 8-4 could each be solutions of equation (8.27) corresponding to different values of k . The number of plausible solutions (choices of k) can be large. Because the dominant term in the series expansion of $\cos^{-1}(x)$ is just x , incrementing k will change ψ_k by approximately λ/pB radians. For the C band SRTM radar (5.3 GHz, $\lambda = 0.0566$ m, $B = 60$ m, $p = 1$) this is $943 \mu\text{rad}$, or about 0.054° . The antenna's elevation beamwidth is on the order of 4° , so approximately 74 different depression angles within the elevation beam are consistent with the wrapped IPD measurement. Consequently, 74 plausible scatterer elevations h_k and ground ranges y_k are consistent with the SAR data at y_0 . This is why our first InSAR method does not produce the actual elevation in practice but instead is ambiguous in elevation.

The second InSAR method also computes ambiguous elevation estimates. For a given ground range y , variations in elevation and variations in IPD are linearly related as shown by equation (8.20). However, if the elevation in a scene varies enough to cause the IPD to change by more than 2π radians, the wrapped IPD will have discontinuities wherever another factor of 2π must be added or subtracted to keep $\tilde{\phi}_{ab}$ in the range $[-\pi, \pi)$. These sudden jumps in the IPD create a *fringe* pattern, as illustrated in Figure 8-7. The amount of elevation variation corresponding to one cycle of the fringe pattern is found by setting

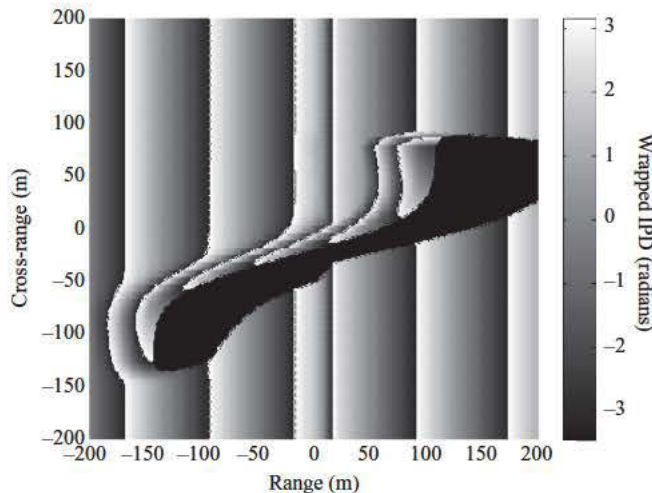


FIGURE 8-7 ■
Wrapped version of
IPD from
Figure 8-6b,
showing just over
five fringe cycles.

$\delta\phi_{ab} = 2\pi$ in equation (8.16). This *fringe height* or *ambiguous height* is

$$h_{amb} = 2\pi |\alpha_{IF}| = \frac{\lambda R \cos \psi}{p B_{\perp}} \quad (8.28)$$

Note that since ψ and thus α_{IF} vary with ground range across the scene, so does the fringe elevation.

Figure 8-7 illustrates the wrapped version of the IPD of Figure 8-6b. For this example $h_{amb} = 26.93$ m in the center of the image. By counting the places where the pattern wraps from almost black back to white, we can see that there are about six and one-third fringes or fringe cycles across the image, consistent with the approximately 40 radians of IPD variation. The height variation in this image is 73.6 m, and $|\alpha_{IF}|$ is about 4.28 at the center and varies only about 10% across the scene. Equation (8.16) then suggests an IPD variation of only 17.2 radians. The remaining variation in IPD is due to the flat earth IPD, which varies 32.2 radians across the scene. (The total IPD variation is less than the sum of these two contributions because the maximum variation due to topography does not occur at the same range as the maximum flat earth IPD.)

If the IPD is not unwrapped, then an InSAR elevation profile obtained using the second InSAR method will produce a wrapped elevation map, with the estimated elevation varying only over a range of h_{amb} meters before wrapping. A similar problem exists with the first InSAR method. Using the wrapped phase implicitly assumes choosing a particular value of k in equation (8.27), restricting the range of depression angles and thus the range of elevations that can be represented. In either case, the restricted range imposed on h is often much smaller than the actual elevation variations to be mapped, so the phase must be unwrapped to produce a useful profile.

8.3.5 Spatial Baseline Effect on Accuracy and Fringe Height

Equation (8.28) shows that the fringe elevation is inversely proportional to the baseline length as projected orthogonal to the LOS, B_{\perp} , which in turn is proportional to the baseline B . A smaller value of B thus allows a larger elevation variation before the IPD wraps. Because it is generally easier and less error-prone to perform phase unwrapping when there are fewer fringes in the IPD, this relationship suggests that smaller values of B are desirable.

On the other hand, equation (8.16) implies that for a given change in elevation, a larger value of B results in a larger change in the IPD. That is, larger baselines enable greater sensitivity in measuring height changes. Conversely, if $\delta\phi_{ab}$ is interpreted as an error in measuring the IPD, then a larger baseline reduces the contribution of the IPD error to the height error. Thus, a trade-off is required in the choice of baseline length between increasing IPD sensitivity to height changes (long baseline), reducing the effect of IPD errors on height estimates (long baseline), and reducing the number of fringe cycles to be unwrapped (short baseline).

8.3.6 Estimating Elevation Relative to a Reference Profile

It is often the case that some information is already known regarding the elevation profile of a scene. For example, a coarse-resolution, low-accuracy (perhaps DTED 1) DEM profile $h_{ref}(x, y)$ of an area may be available, and the intent of the InSAR system is to produce a finer resolution, higher accuracy (perhaps DTED 3) profile estimate $\hat{h}(x, y)$. Define the

differential elevation

$$\begin{aligned}
 \hat{h}_{dif}(x, y) &\equiv \hat{h}(x, y) - h_{ref}(x, y) \\
 &= \alpha_{IF} [\phi_{ab}(x, y) - \phi_{ab}^{ref}(x, y)] \\
 &= \alpha_{IF} \phi_{ab}^{dif}(x, y)
 \end{aligned} \tag{8.29}$$

It is straightforward to compute the IPD $\phi_{ab}^{ref}(x, y)$ that would be observed by our radar when viewing the profile $h_{ref}(x, y)$. First, compute the depression angle $\psi_{ref}(x, y)$ to each pixel using equation (8.2). Next, use that result in equation (8.12) to predict the unwrapped reference IPD $\phi_{ab}^{ref}(x, y)$.

Now suppose our radar is used to newly measure a wrapped IPD $\tilde{\phi}_{ab}(x, y)$ of the same scene. Assuming that $\tilde{\phi}_{ab}(x, y)$ could be unwrapped to obtain $\phi_{ab}(x, y)$, an elevation estimate $\hat{h}_{dif}(x, y)$ could be formed using equation (8.29). However, consider the wrapped version of $\phi_{ab}^{dif}(x, y)$:

$$\begin{aligned}
 \tilde{\phi}_{ab}^{dif}(x, y) &= \langle \phi_{ab}(x, y) - \phi_{ab}^{ref}(x, y) \rangle_{2\pi} \\
 &= \langle \tilde{\phi}_{ab}(x, y) - \tilde{\phi}_{ab}^{ref}(x, y) \rangle_{2\pi} = \langle \tilde{\phi}_{ab}(x, y) - \langle \phi_{ab}^{ref}(x, y) \rangle_{2\pi} \rangle_{2\pi}
 \end{aligned} \tag{8.30}$$

The last equality indicates that the wrapped differential IPD $\tilde{\phi}_{ab}^{dif}$ can be computed by wrapping the difference between the measured IPD, which is inherently wrapped, and the wrapped reference IPD obtained from the reference DEM.

The process of subtracting the wrapped reference DEM IPD from the measured IPD is called interferogram flattening. It removes the gross phase variation, leaving only the differential relative to the reference profile. This tends to greatly reduce the number of fringes (wrapping cycles) in the data, which in turn tends to simplify phase unwrapping. The reference profile is ideally an existing DEM, but the interferogram can be significantly flattened simply by referencing it to a nominal profile such as a flat earth surface (for airborne systems) or a standard Earth ellipsoid such as the World Geodetic System (WGS) 84 model (for spaceborne systems) [19]. In the flat earth case, the required reference DEM IPD is the flat earth IPD of equation (8.19).

Figure 8-8 illustrates the effect of interferogram flattening on the complexity of an IPD. The measured wrapped IPD $\tilde{\phi}_{ab}$ of Figure 8-8a shows approximately 70 rapidly varying fringes. This function could be unwrapped to produce an estimate of ϕ_{ab} , which could then be used in algorithm 1 or 2 to estimate \hat{h} . However, Figure 8-8b shows the result of estimating instead the wrapped differential IPD $\tilde{\phi}_{ab}^{dif}$ by referencing the measured wrapped IPD $\tilde{\phi}_{ab}$ to the WGS 84 smooth surface earth model using equation (8.30). $\tilde{\phi}_{ab}^{dif}$ must still be unwrapped to estimate $\hat{\phi}_{ab}^{dif}$ and then \hat{h}_{dif} and \hat{h} , but now the number of fringes that must be unwrapped is reduced to about 9 or 10. Figure 8-8c references the IPD to an existing DEM of the same area, and the number of fringes is further reduced to about four. Because of this reduction in the number of fringes, it is common to flatten the interferogram before phase unwrapping.

8.3.7 Range Foreshortening and Layover

Standard SAR image formation processing is designed to assume that all echoes originate from a two-dimensional flat surface. Equivalently, the three-dimensional world is projected

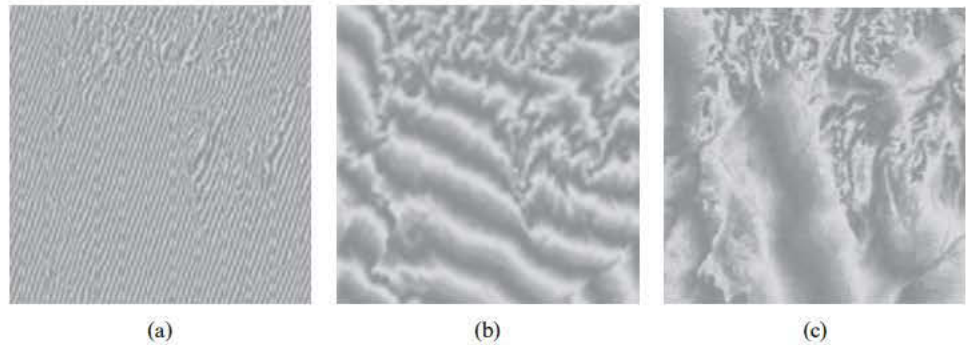


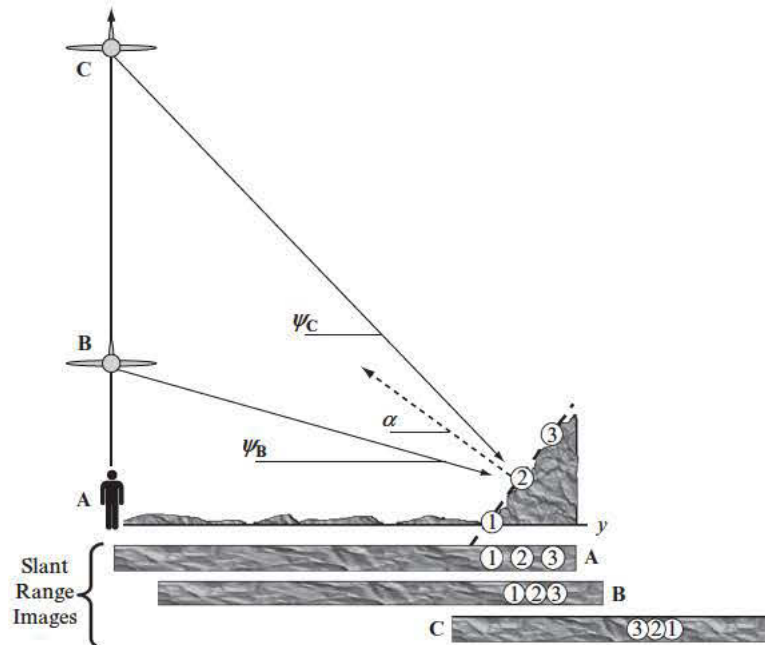
FIGURE 8-8 ■ Interferogram flattening reduces the number of fringes to be unwrapped. (a) Raw interferogram. (b) Flattened to the WGS 84 Earth ellipsoid. (c) Flattened to an existing DEM. (Courtesy of European Space Agency.)

into a two-dimensional plane. As illustrated earlier in Figure 8-4, this means that for a given cross-range coordinate x_0 , the reflectivity observed in the image at ground range coordinate y_0 may be due to a scatterer \mathbf{P}_1 that is actually at ground range y_1 but that is elevated by the appropriate amount h such that it is projected to the position of scatterer \mathbf{P}_0 in the figure. Assuming that $R \gg h$, the isorange contour may be approximated as locally planar at the location of \mathbf{P}_0 and \mathbf{P}_1 and $\delta\psi$ can be assumed small, so that the scatterer projects to ground range

$$y_0 \approx y_1 - h \tan \psi_0 \quad (8.31)$$

The imaging of the elevated scatterer at an incorrect ground range coordinate is termed either layover or foreshortening, depending on the terrain slope and grazing angle and the resultant effect on the image. Figure 8-9 illustrates the difference. Three scatterers are

FIGURE 8-9 ■ Layover and foreshortening. The scene as viewed from aircraft B is subject to foreshortening. When viewed from aircraft C, it is subject to layover.



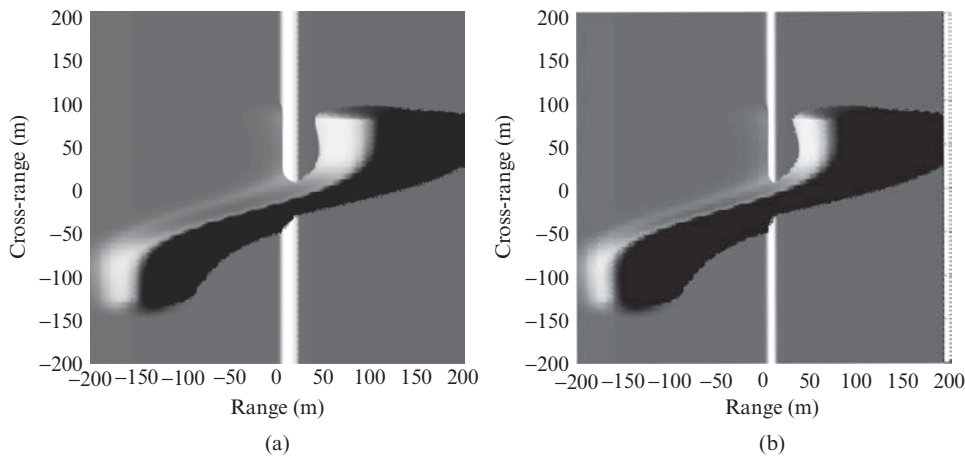


FIGURE 8-10 ■
 (a) Simplified reflectivity image of notional elevation model without foreshortening.
 (b) Same image with foreshortening.

highlighted on sloped terrain. A ground observer and two airborne observers image the scene and project it into the ground plane. The ground observer **A** observes the true ground ranges of the scene. Airborne observer **B** measures the scatterers to be at longer ranges due to the platform altitude. Because the LOS is below the normal to the terrain slope in the vicinity of scatterers 1, 2, and 3, they are imaged as occurring in the correct order, but with their spacing compressed. This compression of range, while maintaining the correct ordering of features, is called foreshortening. Airborne observer **C** images the scene from a higher altitude, with an LOS above the terrain normal. The apparent ranges are longer still, but now the scatterers are imaged in reverse order. That is, a plane wave propagating along the LOS will strike scatterer 3 first because it is actually closer to the radar than scatterer 2, and so forth. The term layover refers to this reversal of apparent range. Layover is particularly evident when imaging vertical walls, such as the sides of buildings in urban areas, where the radar is always above the (horizontal) normal to the wall surface.

In sidelooking operation, the sole case considered here, foreshortening or layover occurs only in the range coordinate. In squinted operation, it occurs in both range and cross-range; details are given in [14,20].

Figure 8-10 illustrates foreshortening on the notional example of Figure 8-6. Figure 8-10a is a simplified image of the scene without accounting for foreshortening. The brightness of a pixel is determined by the relative reflectivity of that pixel, which is modeled with a constant gamma area reflectivity [12,21].⁸ Recall that the radar is located 5 km to the left of the image center and is elevated 2 km. The S-shaped hill has the highest reflectivity (brightest pixels) at its beginning and end, when the hill side is oriented toward the radar and presents a clutter surface closer to normal to the radar line of sight. The reflectivity is lower in the middle of the hill, where the radar line of sight strikes it at more of an angle. The image is also quite bright at the plateau edge, which presents a nearly orthogonal face to the radar LOS. This brightening of facing slopes is quite common in SAR imagery. Of course, the hill also casts a shadow, which is more elongated at the end of the hill due to its greater elevation there.

⁸The effect of range on clutter reflectivity is not accounted for in this simple illustration; the range varies less than 10% across the image.

FIGURE 8-11 ■
SAR image of the
U.S. Capitol showing
layover distortion of
the dome. (Courtesy
of Sandia National
Laboratories.)

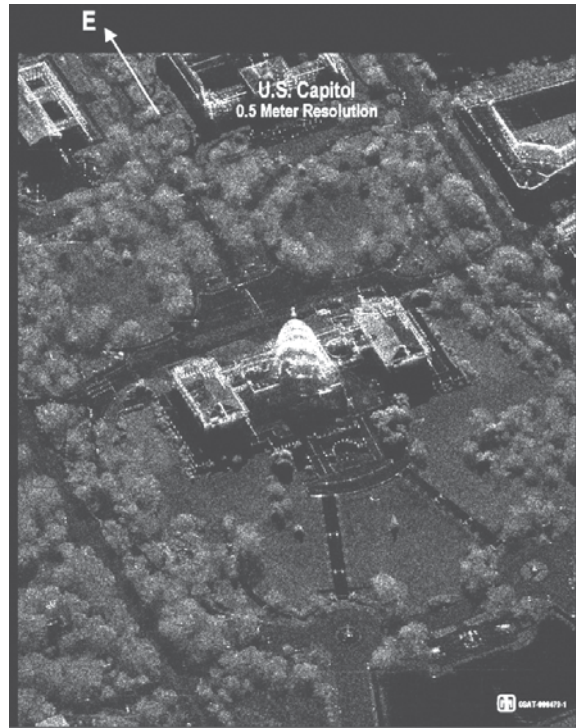


Figure 8-10b is the same scene with foreshortening distortion accounted for. This is the slant range SAR image naturally generated by the radar. Foreshortening of the bright bands is clearly evident in the beginning and end of the hill and in the edge of the plateau as well as in the shortening of the range extent of the image from 400 m to about 392 m, as shown by the end of the data at a range of +192 m.

Figure 8-11 is an airborne image of the U.S. Capitol. The radar is imaging the scene from a position above the top of the image. Layover toward the top of the image is clearly evident in the distorted dome. Figure 8-12a is a satellite SAR image of the volcano Mount Vesuvius in Italy collected by the ERS-2 SAR, again with the radar viewing the scene from a position above the top of the image [22]. Compared with the satellite photo in Figure 8-12b, foreshortening and brightening of the slopes of the volcano facing the radar are both evident.

8.3.8 Other Three-Dimensional Coherent SAR Techniques

The InSAR approach that is the focus of this chapter, and of virtually all operational 3-D SARs, is based on using a two-element interferometer orthogonal to the flight path to estimate angle of arrival and thus scatterer elevation. More advanced approaches using multiple spatial samples orthogonal to the flight path are being investigated for obtaining multiple-resolution cell imagery in the vertical dimension. An introduction to the two major multiple sample methods is given in [23].

The first, called tomographic SAR, uses some number N of vertical apertures instead of the two used in standard InSAR. These could be obtained using multiple flight passes or by building a multiple-aperture system. Analysis of the phase signature observed across the

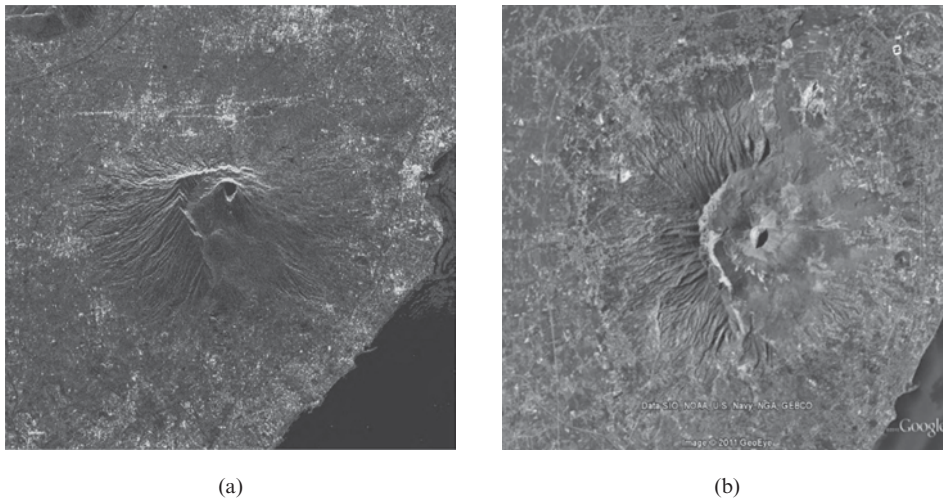


FIGURE 8-12 ■ Images of Mount Vesuvius. (a) ERS-2 SAR image showing foreshortening and brightening of the slopes of the volcano facing the radar (upper slopes in this image). Compare with Google Earth satellite image in (b). (SAR image courtesy of the European Space Agency (ESA). Google Earth image © 2011 Google and © 2011 GeoEye.)

tomographic aperture at a fixed time and slant range for scatterers of different elevations shows that it consists of a term that is linear in scatterer elevation, and a quadratic term that does not depend on scatterer elevation. Applying an appropriate set of weights across the tomographic aperture outputs for a given pixel creates a beam steered to a specified elevation. Varying the weights to vary the focus elevation allows creation of a vertical profile of the scattering scene. The angular resolution obtained is on the order of $\lambda/D_{t\perp}$, where $D_{t\perp}$ is the tomographic aperture size orthogonal to the line of sight. For pixels at a range R , this becomes a vertical resolution on the order of $\lambda R/D_{t\perp}$ meters. As with any sampled aperture, attention is needed to aperture spacing and weight tapering to control grating lobes and sidelobe levels [12].

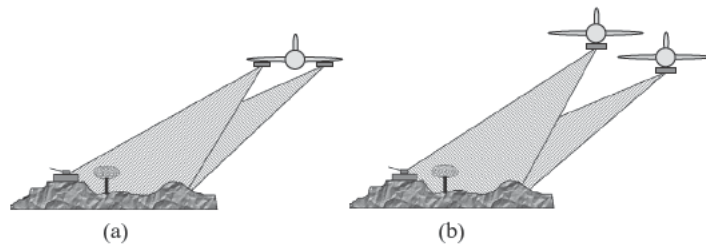
The phase model above leads to an alternative vertical resolving method using Fourier transforms. If the vertical aperture samples are evenly spaced, then the phase slope across the aperture is identical to that of a sampled sinusoid whose frequency depends on the scatterer elevation. A discrete Fourier transform (DFT) of the vertical data then produces a vertical image for a given pixel. In practice, interpolation in the vertical dimension is needed to produce data on an evenly sampled vertical grid prior to the DFT.

8.4 | INSAR OPERATIONAL CONSIDERATIONS

8.4.1 One-Pass versus Repeat-Pass Operation

InSAR data collection operations can be characterized as one-pass or repeat-pass. Figure 8-13 illustrates these two cases. In one-pass processing, a platform with two physical receive apertures, each with an independent coherent receiver, collects all of the radar data needed in a single pass by a scenario of interest. As discussed earlier, either one or two transmitters can be used and the receivers can be displaced horizontally as shown, vertically, or along any baseline orthogonal to the velocity vector. The two SAR images $a(x, y)$

FIGURE 8-13 ■
InSAR data
collection modes.
(a) One-pass.
(b) Repeat-pass.



and $b(x, y)$ are formed from the two receiver outputs. In repeat-pass operation, the platform requires only a conventional radar with a single receiver but makes two flights past the area of interest. The two flight paths must be carefully aligned to establish the desired baseline.

The advantages of one-pass operation are the relative ease of motion compensation and baseline maintenance, since the two apertures are physically coupled, and the absence of any temporal decorrelation of the scene between the two images (see Section 8.6.1.3), since they are collected simultaneously. The major disadvantage is the cost and complexity of the multireceiver sensor. Conversely, the major advantage of repeat-pass operation is the ability to use a conventional single-receiver SAR sensor, while the major disadvantage is the difficulty of controlling the two passes and compensating the data from the two receivers to carefully align the effective collection paths, as well as the likelihood of significant temporal decorrelation of the scene between passes.

Because of the motion compensation difficulty for airborne systems, repeat-pass operation is more easily applied to spaceborne systems, where the two passes are implemented as two orbits of the same spacecraft. The lack of atmospheric turbulence and the stable and well-known orbital paths make it easier to produce an appropriate pair of InSAR images. On the other hand, if different orbits of one satellite are used to establish the baseline, suitable orbits can easily be at least a day apart. For example, in some cases the RADARSAT system has used orbits 24 days apart to form interferometric images [13]. In such systems, temporal decorrelation may be a significant limiting factor. One means of addressing this problem is formation flying of two or more cooperating SAR satellites. Another method requiring only a single satellite sensor is persistent scatterer interferometry. Both are discussed in Section 8.6.1.

8.4.2 Spaceborne versus Airborne InSAR

InSAR DEMs can be generated from both satellite and airborne platforms [24]. Satellite systems such as SRTM, RADARSAT, and TerraSAR-X provide moderate post (elevation sample) spacing, typically 30 to 100 m. Vertical accuracies are on the order of 5 to 50 m. Airborne systems generally generate finer resolution SAR maps, which in turn support closer post spacing and higher accuracy; airborne systems routinely provide vertical accuracies of 1 to 5 m on a post spacing of 3 to 10 m. While achievable SAR resolution is independent of range in principle, practical considerations such as the decrease in SNR⁹ and the increase in required aperture time with increasing range favor shorter ranges for

⁹The “signal” in SAR is most often echo from distributed area clutter, so that it might be more appropriate to use clutter-to-noise ratio. We will use signal-to-noise ratio in this chapter, since this is applicable to SAR pixels dominated by either area clutter or isolated point scatterers.

very fine-resolution SAR. Airborne laser radar altimeter (light detection and ranging, or LIDAR) systems provide the highest quality DEM data, with post spacing of 0.5 to 2 m and vertical accuracy on the order of tens of centimeters [25,26].

Satellite systems provide nearly global coverage at relatively low cost. Their responsiveness and availability depends strongly on when an orbit will provide coverage of the desired region. Numerous concepts have been proposed for satellite constellations that would provide more continuous and rapid global InSAR coverage, but none are yet fielded. Airborne InSAR lacks global coverage capability and has a higher cost per unit area, making it most useful for localized mapping. Timely access to a given region can be limited by airspace restrictions or simply the time required to transport an instrument to the area. The much lower altitude of airborne systems makes the area coverage rate much smaller as well.

Airborne systems require high-precision motion compensation to overcome the defocusing and mislocation effects resulting from path deviations caused by vibration, atmospheric turbulence, and winds. These effects are much reduced or absent in spaceborne systems, although platform orbit and attitude must still be carefully controlled. Spaceborne systems are subject to dispersive ionospheric propagation effects, principally variable path delays in repeat-pass systems up to tens of meters, that are absent in one-pass airborne systems [24]. Both airborne and spaceborne systems suffer potential errors due to differential delay through the wet troposphere. For example, using 1995 Shuttle Imaging Radar-C (SIR-C) repeat-track data, Goldstein [27] estimates root mean square (RMS) path length variations of 0.24 cm at both L and C band. For the baselines used in those experiments, this translates into a 6.7 m RMS elevation estimate error.

8.4.3 Relation to Other Technologies for Estimating Elevation Profiles

A variety of technology alternatives exist for generating high-accuracy, fine-resolution DEMs. In addition to InSAR, these include at least optical and radar photogrammetry and LIDARs. Figure 8-14 illustrates the trade-offs in area coverage versus DEM accuracy for various DEM technologies and sensors [28], while Figure 8-15 shows the approximate relative costs and accuracies of some of these technologies as of the early 2000s. Additional comparisons of these technologies are given in [2,25,26,29,30].

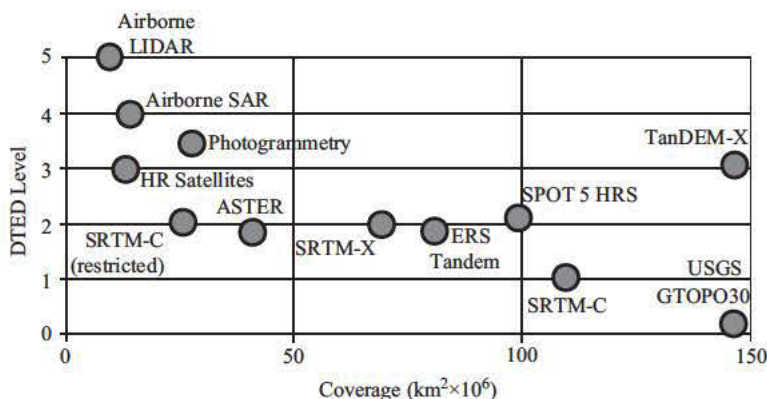
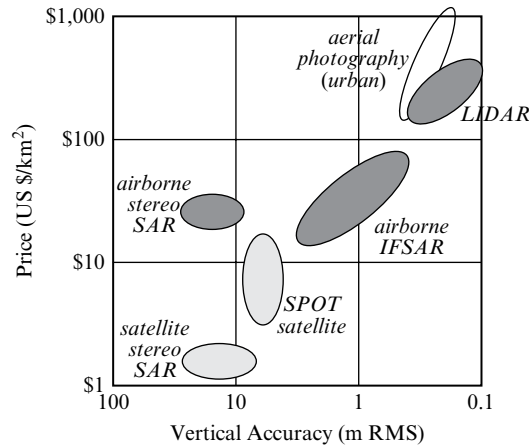


FIGURE 8-14 ■ Relative coverage and accuracy of selected DEM generation technologies. (From Fallor and Weber [28]. With permission.)

FIGURE 8-15 ■
Relative cost and accuracy of DEM generation technologies. (From Mercer [26]. With permission.)



8.5 | INSAR PROCESSING STEPS

Formation of an InSAR image involves the following major steps:¹⁰

- Image pair generation
 - Formation of the two individual SAR images, $a[l, m]$ and $b[l, m]$
 - Registration of the two images
- Estimation of the wrapped interferometric phase difference $\tilde{\phi}_{ab}[l, m]$
 - Formation of the interferogram $I_{ab}[l, m]$
 - Filtering of $I_{ab}[l, m]$ to reduce phase noise
 - Extraction of the wrapped interferometric phase difference $\tilde{\phi}_{ab}[l, m]$ from $I_{ab}[l, m]$
 - Reference DEM phase removal to form $\tilde{\phi}_{ab}^{dif}[l, m]$
- Two-dimensional phase unwrapping to estimate the unwrapped differential phase $\phi_{ab}^{dif}[l, m]$ from $\tilde{\phi}_{ab}^{dif}[l, m]$
- Estimation of the terrain map from the unwrapped differential phase $\tilde{\phi}_{ab}^{dif}[l, m]$
 - Baseline estimation
 - Apply one of the two InSAR algorithms
 - Determine the depression angle to each pixel and then solve for elevation
 - Scale the unwrapped phase map to obtain the relative elevation map $h[l, m]$
 - Orthorectification to develop an accurate three-dimensional map
 - Geocoding to standard coordinates and representations

Each of these steps is now discussed in turn.

¹⁰The indices $[l, m]$ instead of (x, y) are used when we want to emphasize that the various maps have been sampled in range and cross-range.

8.5.1 Image Pair Generation and Registration

The images are formed using any SAR image formation algorithm appropriate to the collection scenario and operational mode, such as the range-Doppler or range-migration (also called ω - k) stripmap SAR algorithms, or polar format spotlight SAR algorithms. Many of the algorithms in common use today are described in Chapters 6 and 7 of this text. More basic introductions to SAR image formation algorithms are given in Chapter 21 of [12] and in [21].

Because the elevation estimation depends on the difference in phase of the echo from each pixel at the two apertures, it is important to ensure that like pixels are compared. The slightly different geometries of the two offset synthetic apertures result in geometric distortion of the images relative to one another, so an image registration procedure is used to warp one complex image to align well with the other. Many registration procedures have been developed in the image processing and photogrammetric literature [30,31]. For one-pass InSAR, the registration of the two images is usually relatively straightforward given the fixed and well-known geometry of the two apertures, although the baseline attitude and orientation must still be determined precisely. In some one-pass systems, the physical structure is subject to significant flexure, vibration, and oscillation, necessitating the use of onboard laser metrology systems to aid in determining the baseline length and orientation. Examples include the 60 m mast used by SRTM and the GeoSAR airborne system, which places P band apertures at the two aircraft wing tips.

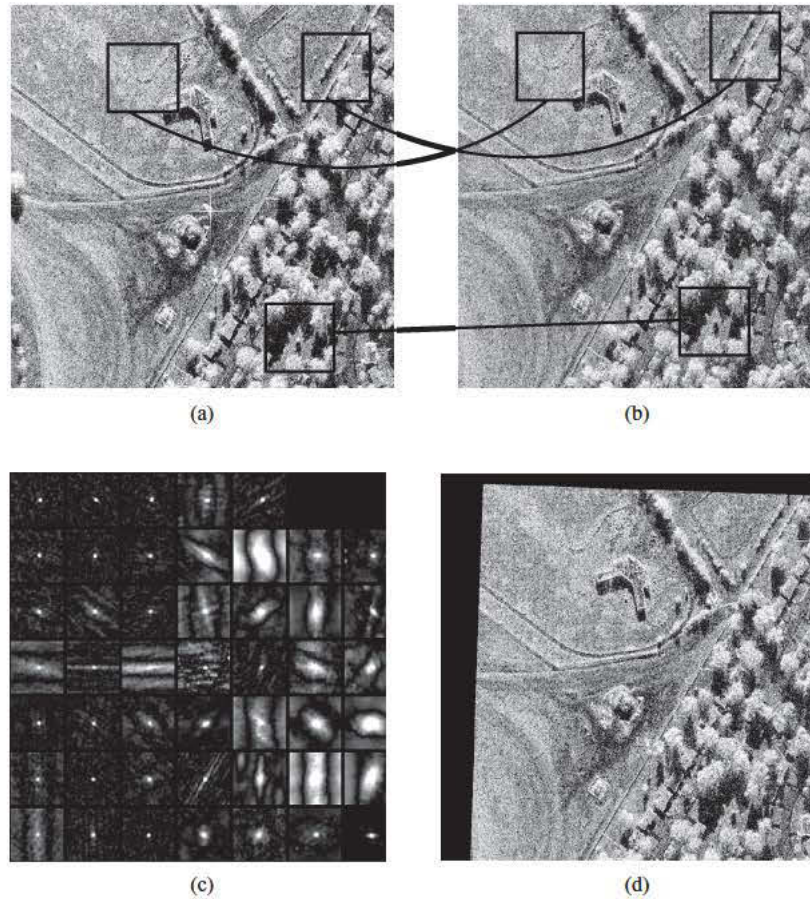
Registration is more difficult in repeat-pass systems, where the baseline can change within the image aperture time and the platform on one pass may be rotated with respect to its attitude on the other pass, creating misregistrations that vary significantly across the scene. It is common to use a two-step procedure, with an initial coarse registration step followed by a fine-grain refinement step.

A common registration approach in InSAR is to use a series of correlations between small subimages of each SAR map to develop a warping function. This concept, called control point mapping or tie point mapping, is illustrated in Figure 8-16. The two InSAR images $a[l,m]$ and $b[l,m]$ are shown in Figure 8-16a and Figure 8-16b, respectively. By examining the images carefully, it can be seen that one is shifted with respect to the other. Take $a[l,m]$ as the master image; $b[l,m]$, now called the slave image, will be warped to align with $a[l,m]$. The same basic approach can be used for both stages, but with possibly different methods of tie point generation, correlation, and warping and interpolation.

The approach begins by identifying a set of tie points in one of the images, and selecting a window of data around each tie point. A data region in the slave image expected to contain the same feature is also identified. A two-dimensional cross-correlation $s_{ab}^n[l,m]$ of each of the corresponding pairs of subimages is formed; the superscript n indicates the n -th subimage. In Figure 8-16, 49 tie points were identified; the data regions around three of them are shown. The magnitude of the 49 resulting correlation functions is shown in Figure 8-16c. For each subimage correlation, if the correlation peak occurs at lag $[0,0]$, it indicates that the two subimages were well aligned in both range and cross-range. If the correlation peak occurs at some nonzero lag $[k_l, k_m]$, it suggests that that region of the secondary image is offset from the corresponding region of the master image by k_l pixels in range and k_m pixels in cross-range.

Many methods are used to identify good tie points (i.e., stable and well localized in two dimensions), including persistent scatterer techniques (see Section 8.6.1.3), thresholding the gradient of the image magnitude and analyzing its curvature [32], and using the

FIGURE 8-16 ■
 Illustration of image registration via tie point mapping.
 (a) Master image with three subregions indicated.
 (b) Slave image with shift of the highlighted regions visible.
 (c) Total of 49 subimage cross-correlation function magnitudes.
 (d) Slave image after warping to align with master image.
 (Courtesy of Sandia National Laboratories.)



intersections of roads and other linear features. For coarse registration, another method is to simply divide the master into relatively large, regularly spaced subregions and then correlate that with the entire slave image to find the highest correlation and identify the best-matching subregion of the slave image.

Registration to within a fraction of a pixel is required for good phase unwrapping results. A typical standard is 1/8 of a resolution cell or better. This standard is based on coherence loss due to misregistration; see Section 8.6.2.2. High-fidelity systems may require estimation to within 1/20 of a pixel or better [33,34]. Consequently, the correlation peak must be located to subpixel accuracy. This can be done by one of several techniques: using oversampled image data, using frequency domain correlation, or using interpolation of the correlation peak. The quality of the subimage correlations also varies significantly, as was seen in Figure 8-16c. If a subregion has low reflectivity, is shadowed, or corresponds to a relatively featureless terrain such as a body of water, the correlation peak is likely to be diffuse and may be a poor indicator of the required warping. Such subimages can be detected by measuring the peak-to-RMS ratio of the subimage correlation:

$$\xi = \frac{\max \{ (s_{ab}^n[l, m])^2 \}}{\sqrt{\frac{1}{LM} \sum_l \sum_m s_{ab}^2[l, m]}} \quad (8.32)$$

ξ takes on higher values for sharp correlation peaks than for diffuse peaks. ξ can then be thresholded to eliminate unreliable tie point candidates.

Now that tie points have been identified and their relative position in each image estimated, the slave image must be transformed to align with the master. A typical approach computes an affine transformation wherein the warping function is a 2-D polynomial that maps tie points in the slave image to the corresponding locations in the master image. A system of equations can be established using the tie point displacements previously found and solved to give the warping function. This procedure is effective at compensating global translations, rotations, skews, and scale changes and works well in relatively flat regions. However, in the second (fine grain) registration stage, it is more common to use a local warping procedure, such as a thin-plate spline method, which can account for nonlinear distortions [32].

In general, the image warping function will require shifting slave image data by fractional pixels, so resampling (interpolation) is required to produce values on the same coordinate grids for both images. Resampling of $b[l, m]$ can be done with any number of interpolation methods. A typical choice that provides adequate quality with relatively low computation for coarse registration is a simple bilinear interpolator. High-fidelity systems may require higher-order interpolators for good results. For fine registration using thin-plate spline methods, the spline is itself the interpolating kernel, so that the two steps of warping and resampling can be combined.

Numerous additional details and extensions for InSAR image registration are described in the literature. Some global skews and translations can be corrected in the original data collection and image formation. A multiresolution, iterative approach may be necessary if misalignments are severe or tie point generation is difficult. Registration techniques have been suggested that use subbanding of the image data to estimate registration errors without cross-correlation or interpolation computations [35]. Despite the additional Fourier transforms required, it is claimed that these techniques can achieve registration accuracies of a few hundredths of a resolution cell with reduced computational complexity.

8.5.2 Estimation of the Wrapped Interferometric Phase Difference

Once the two images are registered, the wrapped phase $\tilde{\phi}_{ab}[l, m]$ must be computed. Because the clutter within a given resolution cell of one image is typically modeled as a random process, the interferogram is also a random process. The maximum likelihood estimator of the wrapped phase map is the phase of an averaged interferogram [20,36]:

$$\tilde{\phi}_{ab}[l, m] = \arg \left\{ \sum_{n=1}^N a[l, m] b^*[l, m] \right\} = \arg \left\{ \sum_{n=1}^N I_{ab}[l, m] \right\} \quad (8.33)$$

The N interferogram samples averaged in this equation can be obtained by dividing the SAR data bandwidth into N subbands, forming degraded-resolution images from each, and averaging the interferograms or by local spatial averaging of a single full-bandwidth interferogram. The latter technique is most commonly used, typically with a 3×3 , 5×5 , or 7×7 window. Error sources limiting the accuracy of IPD estimation are discussed in Section 8.6.

Interferogram flattening is often implemented at this point. In this stage, the wrapped reference DEM or flat earth phase function $\tilde{\phi}_{ab}^{ref}[l, m]$ is subtracted from $\tilde{\phi}_{ab}$ and the result

rewrapped into $[-\pi, \pi)$, giving the wrapped IPD due to terrain variations relative to the reference DEM, $\tilde{\phi}_{ab}^{dif}$. Flattening the interferometric phase reduces the total phase variation, somewhat easing the phase unwrapping step discussed next.

8.5.3 Two-Dimensional Phase Unwrapping

The two-dimensional phase unwrapping step to recover $\phi_{ab}^{dif}[l, m]$ from $\tilde{\phi}_{ab}^{dif}[l, m]$ is the heart of InSAR processing. Unlike many two-dimensional signal processing operations such as fast Fourier transforms (FFTs), two-dimensional phase unwrapping cannot be decomposed into one-dimensional unwrapping operations on the rows and columns. Two-dimensional phase unwrapping is an active research area; a thorough analysis is given in [37], while [38] provides a concise introduction.

Before continuing, it is useful to take note of an inherent limitation of phase unwrapping. Adding some multiple of 2π radians to the entire phase map ϕ_{ab}^{dif} results in the same value for the wrapped IPD $\tilde{\phi}_{ab}^{dif}$. Thus, the unwrapped phase is not unique, and even the best phase unwrapping algorithm can recover the actual IPD only to within a multiple of 2π radians. Phase unwrapping can hope to produce good relative elevation maps, but not absolute elevation maps. Approaches to finding absolute elevation are discussed in Section 8.5.7.

Many traditional phase unwrapping techniques can be classified broadly as either path-following methods or minimum norm methods. Many variants of each general class exist; four algorithms of each type are given in [37]. A newer approach is based on constrained optimization of network flows [39]. Many current systems use variants of these three basic methods. InSAR phase unwrapping on real-world data is an extremely difficult problem due to many factors, including low signal-to-noise ratios, shadow regions, foreshortening, and phase aliasing. These issues and algorithms to deal with them are addressed in detail in [37–41].

New methods of 2-D InSAR phase unwrapping continue to be proposed and evaluated. For instance, techniques to combine the IPD filtering for noise reduction with the unwrapping step using Kalman filters or particle filters are proposed in [42] and [43], respectively, while [44] proposes a wavelet-domain method. Here, we introduce only the most basic concepts of the major classes of two-dimensional phase unwrapping algorithms.

8.5.3.1 Path-Following Method

The path-following approach, which might be better called an integration approach, can be viewed as an extension of one-dimensional phase unwrapping. First, consider a one-dimensional sinusoid of frequency f_0 hertz; the Nyquist sampling rate for this signal is $f_s > 2f_0$. If the sinusoid is sampled at the Nyquist rate or higher, the change in phase from one sample to the next is guaranteed to be less than π radians. Based on this fact, it is well-known that an unaliased one-dimensional wrapped phase signal can be uniquely unwrapped (to within an additive multiple of 2π) by simply starting at one end of the signal and integrating (summing) the wrapped phase differences [18,37].

The path-following approach extends this idea to two dimensions by integrating along an arbitrary path in the two-dimensional discrete $[l, m]$ plane. The difference in phase between any two pixels should not depend on the path taken from one to the other. In practice, however, it can and does. The major reasons for such path dependence include phase noise and pixel-to-pixel phase changes of more than π radians due to aliasing. As a

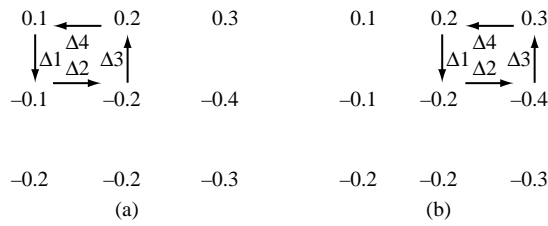


FIGURE 8-17 ■ Illustration of path dependence and residues in two-dimensional phase unwrapping. (a) Path with no residue. (b) Path with residue of -1 . See text for details. (From Ghiglia and Pritt [37]. With permission.)

practical matter, aliasing can be very difficult to avoid: large and sudden changes in actual terrain elevation, say, at cliffs or building sides, can cause large changes in the actual IPD.

Path-dependent data can be recognized by a simple test. Consider the idealized 3×3 segment of wrapped phase data in Figure 8-17. The values shown are in cycles; thus, a value of 0.1 represents a wrapped phase value of 0.2π radians. Because wrapped phases are in the range $[-\pi, +\pi]$, the values in cycles are in the range $[-0.5, +0.5]$. Path dependence can be tested by integrating the wrapped phase difference around a closed path. Because we start and end at the same pixel, the phase values at the beginning and end of the path should be the same. Consequently, the integral of the phase differences around such a path should be zero. In Figure 8-17a, the sum of the differences of the wrapped phase around the path shown is

$$\Delta 1 + \Delta 2 + \Delta 3 + \Delta 4 = (-0.2) + (-0.1) + (+0.4) + (-0.1) = 0 \quad (8.34)$$

However, the path in Figure 8-17b has the sum

$$\Delta 1 + \Delta 2 + \Delta 3 + \Delta 4 = (-0.4) + (-0.2) + (-0.3) + (-0.1) = -1 \quad (8.35)$$

This occurs because $\Delta 3 = +0.7$ is outside of the principal value range of $[-0.5, +0.5]$ and therefore wraps to $0.7 - 1.0 = -0.3$. In this second case, the closed-path summation does not equal zero, indicating an inconsistency in the phase data. A point in the wrapped IPD map where this occurs is called a residue. The particular residue of Figure 8-17b is said to have a negative charge or polarity; positive residues also occur. Conducting this test for each 2×2 pixel closed path is a simple way to identify all residues in the wrapped IPD map. If residues exist, then the unwrapped phase can depend on the path taken through the data, an undesirable condition.

A solution to the residue problem is to connect residues of opposite polarity by paths called branch cuts and then to prohibit integration paths that cross branch cuts. The remaining legal integration paths are guaranteed to contain no pixel-to-pixel phase jumps of more than π radians, so integration yields consistent unwrapping results. In a real data set, there may be many residues and many possible ways to connect them with branch cuts. Thus, the selection of branch cuts becomes the major problem in implementing path following. One of the limitations of path-following methods is that portions of the wrapped phase map having high residue densities can become inaccessible, so no unwrapped phase estimate is generated for these areas and holes are left in the unwrapped phase map [38].

The most widely known path-following approach is the Goldstein-Zebker-Werner (GZW) algorithm [45], which is reasonably fast and works in many cases. Many practical systems begin with the GZW algorithm. For instance, the SRTM mission uses a GZW

FIGURE 8-18 ■
Artificial hill image
for demonstrating
phase unwrapping
performance.

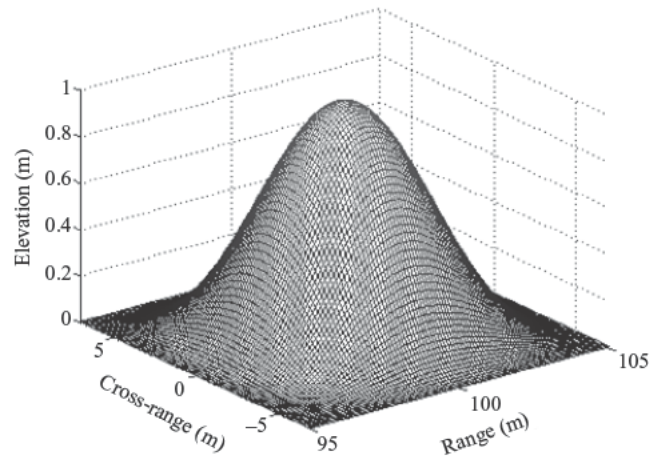
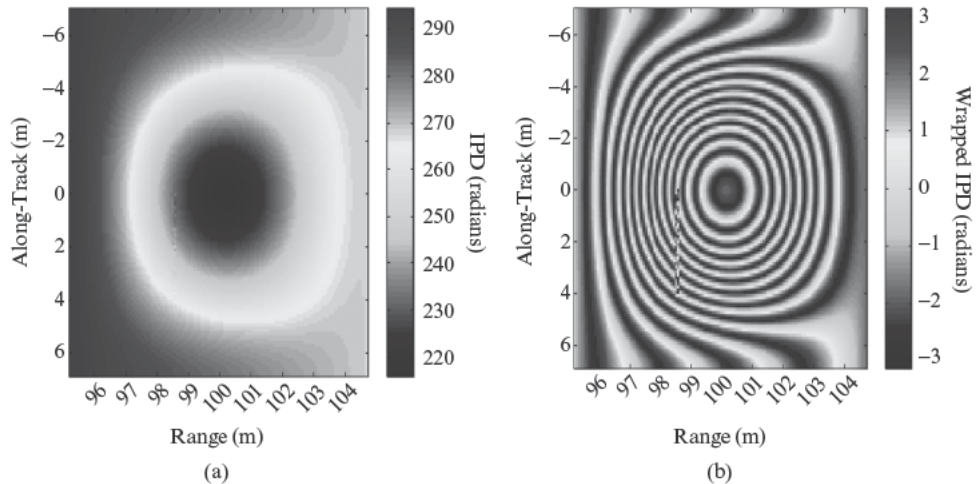


FIGURE 8-19 ■
Interferometric
phase data for the
hill example.
(a) Unwrapped.
(b) Wrapped.



algorithm modified to improve robustness and maximize the unwrappable area [46]. A description of the GZW algorithm is beyond the scope of this chapter; the reader is referred to [37,45] for the details as well as alternative algorithms that can be used when the GZW algorithm fails.

An issue with the GZW algorithm and 2-D phase unwrapping in general is that in real data the number of residues can be large, and the evaluation and selection of branch cuts can become computationally intractable. An extension to the GZW algorithm that improves on speed and completeness while also allowing user weights is given in [40].

As a very simple, idealized example of the path-following approach, consider the hill function shown in Figure 8-18. Figure 8-19a shows the true IPD for this example, while Figure 8-19b shows the wrapped IPD; this would be the starting point for InSAR phase unwrapping.¹¹ Line-of-sight shadowing was not implemented for this simulation. Notice

¹¹This example is from simulations of a synthetic aperture sonar system. See [47] for details.

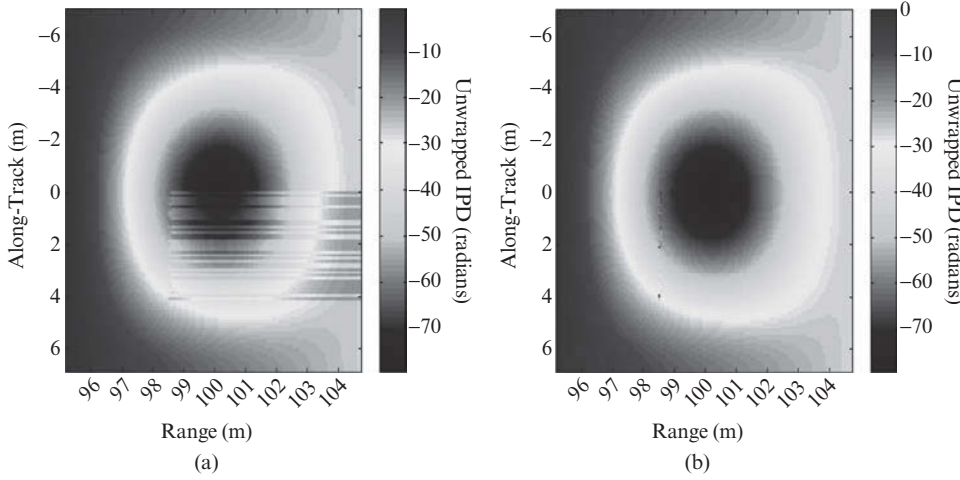


FIGURE 8-20 ■ Phase unwrapping using the path-following technique. (a) Result ignoring residues. (b) Results using the GZW algorithm.

the small linear patch of noisy IPD data at about 98.5 m of range. Such a patch could result from low-reflectivity terrain, shadowing, or data corruption. Applying the residue test on each 2×2 pixel loop in this image would reveal a number of nonzero residues of both positive and negative in the vicinity of this noise patch.

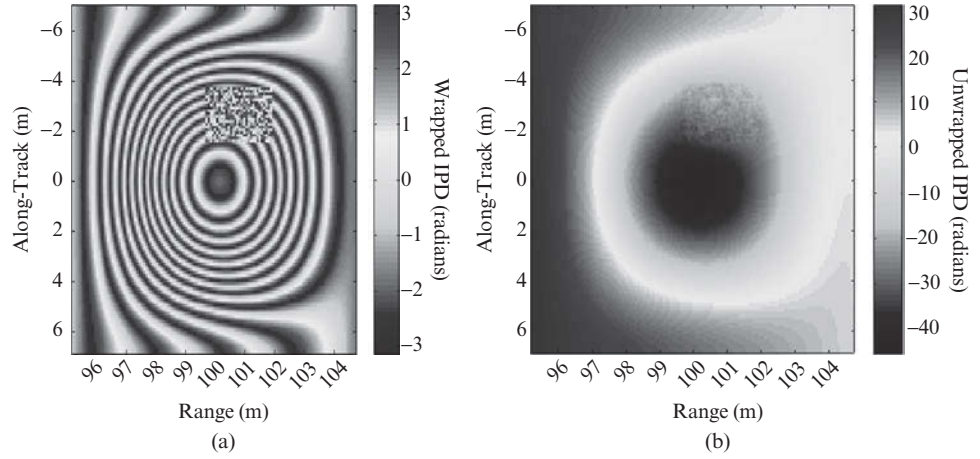
Figure 8-20 shows the result of unwrapping the phase map of Figure 8-19b via path-following techniques. Figure 8-20a is the result obtained with a systematic path that disregards any possible residues. The IPD noise significantly degrades the unwrapped phase. Figure 8-20b used the GZW algorithm to determine branch cuts and then unwrapped along a path that avoided branch cut crossings. In this case, the unwrapped phase is indistinguishable from the original IPD before wrapping except at two small holes at locations that were inaccessible due to the pattern of branch cuts.

8.5.3.2 Least Squares Method

A second major class of two-dimensional phase unwrapping algorithms is least squares methods. The path-following techniques are local in the sense that they determine the unwrapped phase one pixel at a time based on adjacent values. In contrast, the least squares methods are global in the sense that they minimize an error measure over the entire phase map. A classic example of this approach is the Ghiglia-Romero algorithm described in [48]. This technique finds an unwrapped phase function that, when rewrapped, minimizes the mean squared error between the gradient of the rewrapped phase function and the gradient of the original measured wrapped phase. An efficient algorithm exists to solve this problem using the two-dimensional discrete cosine transform (DCT). The simplest version of the algorithm, called the unweighted least squares algorithm, begins by defining the wrapped gradients of the $M \times N$ raw wrapped IPD data:

$$\begin{aligned}\tilde{\Delta}_y[l, m] &= \begin{cases} \langle \tilde{\phi}_{ab}^{dif}[l+1, m] - \tilde{\phi}_{ab}^{dif}[l, m] \rangle_{2\pi}, & 0 \leq l \leq L-2, \quad 0 \leq m \leq M-1 \\ 0, & \text{otherwise} \end{cases} \\ \tilde{\Delta}_x[l, m] &= \begin{cases} \langle \tilde{\phi}_{ab}^{dif}[l, m+1] - \tilde{\phi}_{ab}^{dif}[l, m] \rangle_{2\pi}, & 0 \leq l \leq L-1, \quad 0 \leq m \leq M-2 \\ 0, & \text{otherwise} \end{cases}\end{aligned}\quad (8.36)$$

FIGURE 8-21 ■ Phase unwrapping using a least squares method. (a) Wrapped phase of hill with square noise patch. (b) Unwrapped phase using the unweighted least squares Ghiglia-Pritt algorithm.



These are then combined into a driving function $d[l, m]$:

$$d[l, m] = (\tilde{\Delta}_y[l, m] - \tilde{\Delta}_y[l - 1, m]) + (\tilde{\Delta}_x[l, m] - \tilde{\Delta}_x[l, m - 1]) \quad (8.37)$$

Let $D[q, r]$ be the $M \times N$ two-dimensional DCT_2 of the driving function.¹² The estimate of the unwrapped phase is then obtained as the inverse DCT_2 of a filtered DCT_2 spectrum:

$$\hat{\phi}_{ab}^{dif}[l, m] = \text{DCT}_2^{-1} \left\{ \frac{D[q, r]}{2 \left\{ \cos \left(\frac{\pi q}{M} \right) + \cos \left(\frac{\pi r}{N} \right) - 2 \right\}} \right\} - 1 \quad (8.38)$$

This phase function is then used to estimate the terrain elevation map $h[l, m]$. Note that the DCT domain filter transfer function is undefined for the “DC value” $q = r = 0$, emphasizing again that the overall phase offset of the estimated map $\hat{\phi}_{ab}^{dif}$ is indeterminate. Notice that the least squares IPD estimate is offset from the GZW estimate by about 30 radians.

Figure 8-21a is the wrapped phase for the hill example but with a larger square noise patch added to simulate a low reflectivity or degraded area. Straightforward application of equations (8.36)–(8.38) produces the unwrapped interferometric phase map estimate of Figure 8-21b. The phase noise remains in the unwrapped map. While it appears to have remained localized, in fact it tends to have a somewhat regional influence. This can be seen by closely comparing Figure 8-21b to Figure 8-20b, particularly in the northeast corner. The general smoothing behavior of least squares methods, coupled with their inability to ignore outliers and other corrupted data, means that data errors tend to have a global influence. It can also be shown that these methods tend to underestimate large-scale phase slopes [38]. On the other hand, they require no branch cut or path computations and consequently produce an unwrapped phase estimate everywhere in the map.

The least squares approach lends itself naturally to an extension that incorporates weights on the data. Generally, the weights are related to an estimate of data quality, so high-quality data regions have more influence on the solution than low-quality regions.

¹²There are multiple forms of the DCT in common use. The notation DCT_2 refers to the specific version identified as the DCT-2 in [49].

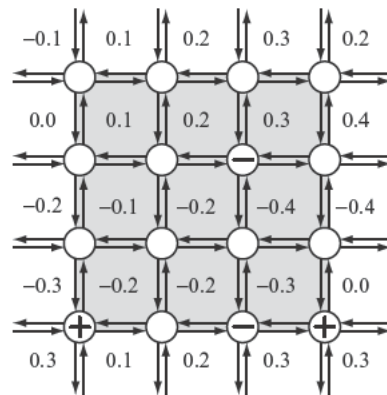


FIGURE 8-22 ■
Network
representation of a
wrapped phase
map. The shaded
area is the same
data shown in
Figure 8-17b.

For instance, a weight matrix for the data of Figure 8-21 would probably place lower weights in the noise region (provided it can be identified) and higher weights elsewhere. Similarly, areas corresponding to shadow regions of a SAR image would carry zero or low weights. The weighted least squares approach does not lend itself so easily to the use of fast transform techniques. Instead, it is typically formulated as the solution of a set of linear equations that are solved by one of a number of iterative algorithms. A more detailed description of these methods is available in [37].

Continuing the earlier example of Figures 8-6 and 8-7, shadowing was simulated and noise was added to the phase data for each receiver prior to IPD formation at an SNR of 20 dB. The noisy wrapped IPD was unwrapped using both the GZW and weighted least squares algorithms, and then the DEM was reconstructed in the unshadowed regions using both InSAR DEM algorithms.¹³ The maximum height error in the unshadowed region using GZW phase unwrapping was about 2.2 m for either InSAR DEM algorithm, and about 2.1 m for least squares unwrapping and either algorithm.

8.5.3.3 Network Flow Method

Constantini [39] described a new approach to phase unwrapping called the network programming or network flow method. This approach represents the gradient of the wrapped IPD map $\tilde{\phi}_{ab}^{dif}[l,m]$ as a constrained network with an error in each value that is an integer multiple of 2π . The unwrapping problem is then posed as a global minimization problem with integer variables. The network structure of the problem allows the application of efficient solution algorithms. The basic approach has been extended and further analyzed, and an excellent description is given in [41].

The network equivalent of the 2-D wrapped phase example of Figure 8-17b is shown in Figure 8-22, extended to a larger area. The numbers between the circular nodes are the wrapped phase values, again in cycles. Each node represents the integrated gradient of the four surrounding phase values. Empty nodes represent a zero residue (no phase inconsistency), while + and - signs inside a node represent residues of +1 and -1. The gray shaded area is the portion of the data that was shown in Figure 8-17b. Note that three additional residues occur along the bottom row of nodes in this extended patch of data. The directed lines connecting nodes represent the phase gradients from one pixel to the next. The flow on a line is the difference in cycles between the unwrapped and wrapped phase

¹³A MATLAB code implementing this example is available online (<http://www.scitechpub.com/pomr2>).

gradient at the two pixels connected by that line, which must be an integer. To remain consistent with the data, the net flow out of a node must equal the residue at that node. The phase unwrapping problem is now equivalent to finding a set of integers describing the flow on each arc.

The solution is not unique. For example, one can simply add one cycle to the flow on each arc of any valid solution to create another valid solution; this corresponds to simply adding a constant offset of 2π radians to each unwrapped phase value. Thus some optimization criterion is needed to choose one particular solution. The minimum cost flow (MCF) algorithm solves this problem by minimizing the total number of extra gradient cycles added to the phase map. Efficient algorithms exist for solving this optimization problem. The network programming approach with the MCF algorithm is generally considered to provide an effective combination of accuracy and efficiency [41].

One issue with the MCF algorithm is that it is relatively memory intensive. As SAR resolutions improve, the size of the data sets that must be unwrapped grows rapidly. Typical approaches for unwrapping large data sets involve partitioning of the data into blocks, solving the smaller subproblems, and then combining the subblock solutions in a way that enforces consistency across block boundaries. One recent variation on this idea performs a whole-scene simplified approximate unwrapping and then refines the solution on a subblock basis [50].

8.5.4 Differences and Commonality of 2-D Phase Unwrapping Methods

The three major classes of phase unwrapping algorithms differ fundamentally in their approach to constructing a phase map consistent with the wrapped IPD. The MCF network flow algorithm minimizes the total number of extra gradient cycles added to the phase map. Path-following algorithms minimize the number of places where the wrapped and unwrapped phase gradients disagree, regardless of the amount of the difference. Least squares methods tend to tolerate many small differences while minimizing large differences, thus allowing small unwrapping errors to persist throughout a scene. Another difference is that any errors persisting in the output of the path-following and network programming methods will be multiples of 2π radians, while with least squares methods errors can take on any value.

In recent years it has been realized that the three major unwrapping algorithm classes previously described can all be considered as special cases of a particular error minimization problem [40]. Consider the following objective function:

$$\Phi \equiv \sum_{l,m} \left\{ w_x[l,m] \cdot |\Delta_x[l,m] - \tilde{\Delta}_x[l,m]|^P + \sum_{l,m} w_y[l,m] \cdot |\Delta_y[l,m] - \tilde{\Delta}_y[l,m]|^P \right\} \quad (8.39)$$

where $\tilde{\Delta}_x$ and $\tilde{\Delta}_y$ are the measured, wrapped IPD gradients, Δ_x and Δ_y are the true IPD gradients, and allowance has been made for including weights w_x and w_y in the solution. Each of the methods discussed corresponds to minimizing Φ with respect to a different L^P norm (i.e., a different choice of P). As discussed already, the least squares method corresponds to choosing $p = 2$. It can be shown that the network flow methods correspond to choosing $p = 1$, while the path-following methods correspond to choosing $p = 0$.

While it can be argued that the goal of phase unwrapping should be to find the L^0 solution, it can also be shown that the L^0 solution of the phase unwrapping problem is NP-hard, in effect stating that it is theoretically intractable. This merely means that unwrapping algorithms should focus on developing efficient heuristic approximations to the solution. This observation also motivates the use of L^1 solutions, since it is well-known in optimization that L^0 and L^1 solutions are often similar but that L^1 solutions can usually be found much more efficiently.

8.5.5 Multibaseline InSAR

An alternative approach to resolving phase ambiguities uses a three phase-center system to provide two different interferometric baselines and therefore two different ambiguity intervals. This approach is very similar in concept to the multiple pulse repetition interval (PRI) techniques commonly used to resolve range and Doppler ambiguities in radar [12]. The longer baseline provides precise but ambiguous elevation estimates, as discussed earlier, while the short baseline, while imprecise, provides a large ambiguous elevation interval that can be used to resolve ambiguities. A particular implementation in the RTV system uses two antennas, one a conventional antenna and the other an amplitude monopulse antenna [7]. A baseline of 0.33 m is formed between the conventional antenna and the monopulse sum port, while a second short effective baseline of 0.038 m is formed by the elevation monopulse antenna. With an RF of 16.7 GHz, $p = 1$, $\beta = 90^\circ$, and nominal range of 5 km and depression angle of 30° , the 0.33 m baseline has an ambiguous elevation interval h_{amb} of about 363 m (equation (8.28)), while the short 0.038 m baseline gives h_{amb} of about 3.15 km, more than enough to represent any expected elevation variation for most terrains. This system requires no phase unwrapping algorithm at all; a very simple algorithm suffices to remove the phase ambiguity in the conventional InSAR image using the short-baseline monopulse angle of arrival information. An additional advantage is that the IPD ambiguity removal is performed on a pixel-by-pixel basis, so errors, if they occur, do not propagate throughout the scene. The cost of this improvement is that three receiver channels and image formers must be implemented, one each for the monopulse sum, monopulse elevation difference, and second antenna channels. Figure 8-23 gives both an orthorectified SAR image of the Pentagon and the corresponding DTED level 4 DEM generated by the RTV system.

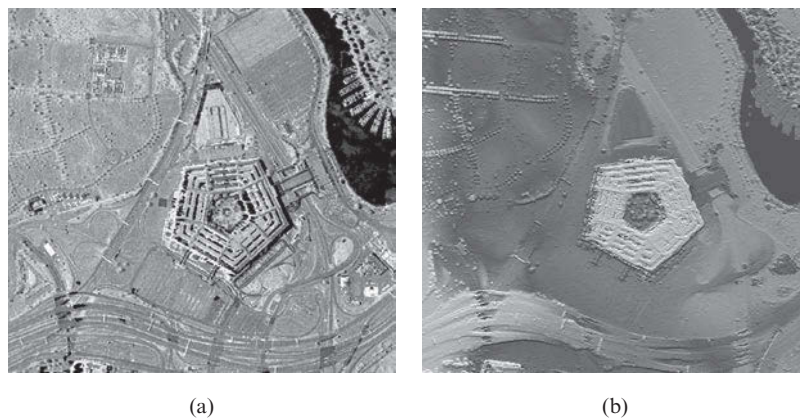


FIGURE 8-23 ■
DEM of the Pentagon generated by the RTV system. (a) Orthorectified SAR image, 0.75 m posts. (b) DTED level 4 digital elevation map. (Courtesy of Sandia National Laboratories.)

8.5.6 Estimation of the Terrain Map from the Unwrapped Phase

Equation (8.16) shows that it is necessary to know the InSAR baseline precisely to accurately scale the now-unwrapped IPD to elevation variations. Typical requirements are that the baseline be known to within a factor of 10^{-4} to 10^{-6} of the absolute slant range to the imaged area. The combination of a relatively rigid mechanical baseline structure with modern GPS and INS data will usually allow the baseline length and orientation to be specified accurately enough in one-pass systems, although in systems with flexible baselines laser metrology may also be needed. However, in repeat-pass systems, navigational data are often inadequate to estimate the actual baseline to the accuracy required. In this case, there is an extra processing step called baseline estimation necessary to estimate the actual baseline length and orientation over the course of the synthetic aperture. A typical approach computes a least squares estimate of the baseline parameters using a series of tie point displacements between intensity images [51].

8.5.7 Absolute Elevation Estimation

Once the baseline has been estimated, the unwrapped phase map $\hat{\phi}_{ab}^{dif}$ is used in one of the two InSAR algorithms to determine the estimated relative elevation profile $\hat{h}(x, y)$. Absolute elevation can be estimated by a variety of techniques. Perhaps the easiest is simply the use of one or more surveyed reference points within an image; elevation relative to this point is then easily converted to absolute elevation. This technique is common for InSAR system test and calibration.

A variation on this idea can be applied when a landmark with known latitude and longitude is contained in the image. That information can be used to compute the ground range y at which the landmark should appear in the SAR image. Equations (8.14) and (8.27) can then be used to determine the number k of 2π phase cycles that should be added to the wrapped IPD to get the absolute IPD consistent with the known ground range, which in turn allows estimation of the absolute elevation.

If surveyed reference points are not available, another method compares the unwrapped relative elevation profile to a lower-resolution existing DEM of the area of interest. An integer multiple k of 2π is added to the entire unwrapped IPD of the area and the elevation profile estimate regenerated. The value of k that produces the smallest RMS residual between the estimated and reference DEM profiles is taken as the correct value, and the associated elevation profile is taken as the absolute elevation estimate. This technique has been used in the SRTM processor [46].

An alternative is to attempt to estimate the correct absolute phase shift, and ultimately absolute elevation, directly from the radar data. At least two methods have been suggested; details of both are given in [52]. The first method splits the fast-time bandwidth of the SAR data in half, producing two image data sets having different effective radar carrier frequencies. SAR image formation and the InSAR processing are completed separately for each half of the data. It can be shown that a differential interferogram formed from the two individual interferograms is equivalent to an interferogram formed using a carrier frequency that is the difference of the two individual half-band center frequencies. This small effective frequency implies a large effective wavelength and thus a large h_{amb} . The individual frequencies can be chosen such that the differential IPD is always in the range $[-\pi, +\pi)$ and is therefore unambiguous. The unwrapped absolute phase can then be estimated from the unwrapped IPD developed earlier and the differential interferogram phase.

The second method relies on using the unwrapped IPD to estimate delay differences between the two InSAR channels. However, these differences must be estimated to a precision equivalent to 1% to 0.1% of a pixel in range, requiring very precise interpolation and delay estimation algorithms. In addition, the technique is sensitive to a variety of systematic errors as well as to phase noise. Accuracy is improved by increasing the interpolation ratio (to support finer cross-correlation peak location estimates) and the degree of spatial averaging of the interferogram (to reduce noise). The accuracy of this method is described in [34].

8.5.8 Orthorectification and Geocoding

The next step in InSAR processing is *orthorectification*, which uses the newly gained elevation information to correct the displacement of image pixels due to foreshortening or layover. For each pixel in the measured (and distorted) SAR image $a(x, y)$, the corresponding estimated elevation $\hat{h}(x, y)$ is used to estimate the foreshortening $-h \tan \psi$ present in that pixel, and a corrected image $a'(x, y)$ is formed by moving the image pixel:

$$a'(x, y + h(x, y) \tan \psi) = a(x, y) \quad (8.40)$$

In general, this involves fractional shifts of the range coordinate, requiring interpolation of the image in the range dimension. If the radar is operated in a squint mode, there is also foreshortening in the cross-range (x) dimension, and a similar shift in the x coordinate is required [20].

It is possible to detect regions where the foreshortening is severe enough to generate actual layover of pixels. As the ground range coordinate y is increased in the uncorrected image, the corrected ground range $y' \equiv y + \hat{h}(x, y) \tan \psi$ is computed. Any region in which the gradient dy'/dy is negative is exhibiting layover [53].

The orthorectified image, along with the corresponding elevation map, locates each pixel in a three-dimensional coordinate system relative to the SAR platform trajectory. To form the final DEM, the data may then be translated to a standard geographical coordinate system or projection, a process known as geocoding [33]. The first step is typically to express the coordinates in the universal Cartesian reference system, which has its origin at the earth's center, the z axis oriented to the north, and the x - y plane in the equatorial plane. The x axis crosses the Greenwich meridian. The next step expresses the elevations relative to an Earth ellipsoid, usually the WGS 84 standard ellipsoid. At this stage, the (x, y, z) coordinates are expressed as a new geographic coordinate set $(\theta_{long}, \theta_{lat}, z)$, where θ_{long} is longitude, and θ_{lat} is latitude. The last step projects the geographic coordinates onto a standard cartographic map, such as the universal transverse Mercator (UTM) projection, which represents points in a north-east-elevation (N, E, z) system. Finally, the data are regridded (interpolated) to uniform spacing in the north and east coordinates.

8.6 | ERROR SOURCES

Since the relative elevation is estimated as a multiple of the interferometric phase difference, errors in the IPD measurements will translate directly into errors in the elevation estimate. These errors arise from several sources, including thermal noise; various processing artifacts such as quantization noise, point spread response sidelobes, and focusing

errors; and decorrelation of echoes between images. The error sources can be broadly characterized as statistical or systematic. Statistical errors introduce a random component to the IPD and thus to the elevation estimate in a DEM. The specific effect on each individual elevation post measurement varies, but the general nature of the error can be evaluated by considering the variance of the elevation. In contrast, systematic errors are nonrandom in nature and affect each pixel deterministically. In addition to being of direct concern, elevation estimation errors also produce orthorectification errors via foreshortening estimation errors. Since elevation errors may contain both systematic and random contributions, so may the foreshortening errors.

8.6.1 Coherence and Statistical Errors

So far, we have implicitly assumed that the complex reflectivity ρ of a resolution cell is a constant. For imaging terrain, it is more realistic to model the reflectivity of a resolution cell as the superposition of the reflectivities of a large number of uncorrelated scatterers of varying radar cross section randomly dispersed through the resolution cell. The complex reflectivity that results from the constructive and destructive interference of scatterers is therefore a random variable, typically with a uniform random phase over $[0, 2\pi)$ radians and an amplitude probability density function (PDF) that is strongly dependent on the type of clutter observed but is often modeled as a Rayleigh PDF [12,21]. The complex interferogram $I_{ab}[l, m]$ is therefore also a random process. Additive noise also affects the SAR image pixels and is reflected in the interferogram.

The fundamental measured quantity in InSAR is the IPD $\phi_{ab}[l, m]$, generally obtained as the argument of $I_{ab}[l, m]$.¹⁴ A measure of the quality of the IPD at a given pixel is the coherence $\gamma[l, m]$, defined for each pixel as

$$\gamma = \frac{\mathbf{E}\{a \cdot b^*\}}{\sqrt{\mathbf{E}\{|a|^2\} \mathbf{E}\{|b|^2\}}} \quad (8.41)$$

where a and b are the two individual SAR images and the $[l, m]$ arguments have been dropped for brevity. In the absence of statistical errors, $|\gamma| = 1$ and $\arg\{\gamma\} = \phi_{ab}$. However, the actual coherence is a random variable due to random variations of pixel reflectivity, additive noise, baseline decorrelation, and temporal decorrelation.

It has been shown that the coherence can be factored into the product [22,54]

$$\gamma = \gamma_{noise} \cdot \gamma_{spatial} \cdot \gamma_{temporal} \quad (8.42)$$

Some authors extend this approach to include additional terms. The following subsections discuss each of these contributions.

8.6.1.1 Thermal and Quantization Noise

The effect of additive thermal noise with an SNR χ in a given pixel is to reduce the coherence of that pixel by the factor [54]

$$\gamma_{noise} = \frac{1}{1 + \chi^{-1}} \quad (8.43)$$

¹⁴In this section we ignore phase wrapping for simplicity. The results are not materially affected.

If sufficient bits are used in the analog-to-digital converters (ADCs) to allow modeling of the quantization noise as white, then χ includes both the thermal and quantization noise. Note that as $\chi \rightarrow \infty$, $\gamma_{noise} \rightarrow 1$. Typical SAR SNRs are 10 to 20 dB, giving γ_{noise} values of 0.9 to 0.99.

8.6.1.2 Baseline Decorrelation

The InSAR baseline causes the two apertures to view a given pixel from slightly different grazing angles. If the baseline and thus the variation in grazing angle $\delta\psi_B$ is great enough, the reflectivity of corresponding pixels in the two SAR images will decorrelate due to the variation of reflectivity with aspect angle. A simple model for the minimum angular change required to decorrelate the radar cross section of a many-scatterer target is given in [21] as

$$\delta\psi = \frac{c}{2Lf} = \frac{\lambda}{2L} \quad (8.44)$$

where L is interpreted as the width of the target perpendicular to the line of sight. This applies to a monostatic radar configuration that is equivalent to $p = 2$ in our InSAR analysis. A similar analysis for $p = 1$ would produce $\delta\psi = \lambda/L$, so the generalization $\delta\psi = \lambda/pL$ covers both cases. Considering the target to be a single resolution cell of ground range extent Δy gives $L = \Delta y \cdot \cos \psi$, where ψ is the nominal grazing angle. Using $R \cdot \delta\psi = B_{c\perp}$, an estimate for the critical baseline perpendicular to the line of sight of the radar is [51]

$$B_{c\perp} = \frac{\lambda R}{p \cdot \Delta y \cdot \tan \psi} \quad (8.45)$$

If the InSAR perpendicular baseline B_\perp is larger than this value, the two images will decorrelate. The loss of coherence due to baseline decorrelation is given by [36]

$$|\gamma_{spatial}| = 1 - \frac{B_\perp}{B_{c\perp}} \quad (8.46)$$

with a minimum value of zero.

Equation (8.46) shows that we require $B_\perp \ll B_{c\perp}$ to avoid significant spatial coherence loss. Continuing the earlier SRTM example with $\Delta R = 15$ m, $p = 1$, and assuming that the baseline is approximately orthogonal to the radar LOS gives a critical baseline of $B_{c\perp} = 1281$ m, far greater than the actual 60 m baseline. For this case, $\gamma_{spatial} = 0.953$, very near the maximum of 1. Thus, SRTM data collection was not expected to suffer significant baseline decorrelation. As another example, the K_u band (16.7 GHz, $\lambda = 0.018$ m) RTV system uses a vertical baseline and also has $p = 1$. In its DTED level 4 mode, it uses a range resolution of 0.3 m at a grazing angle of 45° . Assuming a slant range of 10 km, this gives a critical vertical baseline of 849 m, again far greater than the actual baseline of 0.33 m; $\gamma_{spatial} = 0.996$ in this case. An early example with a smaller safety margin is repeat-pass InSAR processing using the SEASAT satellite. This L band ($\lambda = 0.24$ m) system operates from an 800 km orbit altitude with a steep grazing angle of about 67° and a ground range resolution of about 25 m. Applying the horizontal baseline formula with $p = 2$ estimates the critical baseline at 4532 m. Actual SEASAT baselines formed from orbit pairs viewing the same terrain over a two-week period range from as little

as 50 m to 1100 m. In this case, $\gamma_{spatial}$ is as low as 0.757. Additional analysis and experimental decorrelation data are available in [54].

Another benefit of a large critical baseline is that it permits a larger actual baseline without suffering significant baseline decorrelation. As was seen in Section 8.3.5, a large baseline improves the height measurement sensitivity.

8.6.1.3 Temporal Decorrelation

Temporal decorrelation occurs when there are physical changes in the scattering surface between the two SAR data collections used to form the pair of images. While the time lapse between SAR images is a nonissue in one-pass operation, it can be significant in repeat-pass operation. Pairs of orbital tracks that provide a suitable baseline over a desired portion of the earth's surface can occur days or weeks apart, giving ample time for the surface to change due to weather, man-made disturbances such as construction or vehicle movement, natural disturbances such as subsidence, glacial movement, or earthquakes, and similar effects. Decorrelation time scales as observed from spaceborne systems are typically on the order of several days [16].

Bamler and Hartl [36] give a model for $\gamma_{temporal}$ and report that water surfaces decorrelate in milliseconds so that $\gamma_{temporal} \approx 0$ over water in any repeat-pass InSAR system, while forests tend to have $\gamma_{temporal}$ in the vicinity of 0.2 at C band due to leaf and branch movement between passes. While temporal decorrelation can be used to positive effect for terrain motion mapping and change detection (see Section 8.8) as well as classification of scatterer types, it is an undesirable effect for elevation profile mapping.

One technique for combating temporal decorrelation is reliance on persistent scatterers, which are scatterers in a scene that are exceptionally stable over very long periods of time. Persistent scatterers are typically found in urban areas and rocky terrain. Once identified, InSAR processing can be applied to develop reliable elevation estimates at those points or to perform terrain motion mapping (see Section 8.8.1). Elevation accuracies of ones of meters using images spanning many years have been reported [55].

Another approach avoids temporal decorrelation altogether by using two collection radars flying in close formation, effectively creating the equivalent of one-pass operation. The obvious cost of this approach is the need for two radar satellites. Figure 8-24 illustrates one configuration, the TanDEM-X mission based on the TerraSAR-X radar system [56]. This system uses a pair of orbits that move in a helical relationship to one another, resulting in a InSAR baseline that varies in length and orientation (value of B and β) at different points in the orbit, as shown in Figure 8-24b. Adjustment of the orbit spacing using onboard propulsion units allows fine tuning of the baseline length over a range of a few hundred meters to several kilometers. A dual-frequency GPS is used for fine orbit determination and PRF synchronization, and an X band intersatellite data link provides phase synchronization of the two transmitters.

8.6.1.4 Atmospheric and Weather Effects

Atmospheric conditions can affect estimation of the IPD in several ways. Changes in water vapor content and distribution in the troposphere can lead to variations in signal propagation delays through the atmosphere and thus in radar phase measurements. Since repeat-pass systems may combine measurements taken days apart or longer, the delays may change between acquisitions and the resulting delay differences will introduce errors into the IPD [57]. As discussed in Section 8.4.1, this effect can produce significant DEM

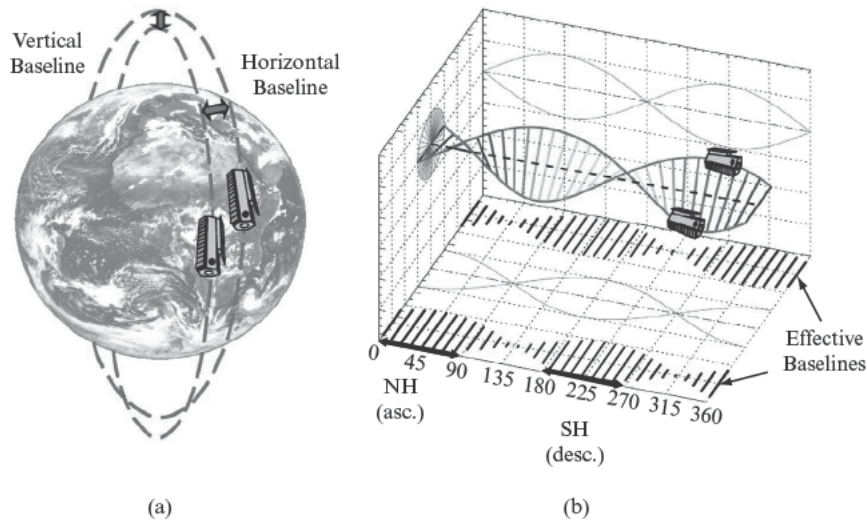


FIGURE 8-24 ■ Formation flying to combat temporal decorrelation. (a) Operational concept for TanDEM-X TerraSAR mission. (b) Variation of the horizontal and vertical baselines during an orbit. (From Krieger [56]. With permission.)

errors in both airborne and spaceborne systems. Numerous papers in the literature explore means of using atmospheric data from ground-based and satellite sensor networks to estimate and compensate these effects. Spaceborne repeat-pass systems are also subject to similar effects due to fluctuating ionospheric delays that can reach tens of meters.

Data collection through a rainy atmosphere degrades the noise component of the IPD coherence by decreasing the signal-to-interference ratio two ways [58]. First, the rain increases atmospheric attenuation, decreasing the received signal strength. At the same time, the radar receives backscattered clutter from the rain, increasing the interference strength. These effects are present in both single- and repeat-pass interferometry.

8.6.1.5 Interferogram Phase Statistics and Multilook Averaging

The effect of reduced coherence on elevation estimation quality can be assessed by relating the coherence to the standard deviation of the interferometric phase, which in turn maps into the standard deviation of the elevation estimate via equation (8.16):

$$\sigma_h \approx \left| \frac{\lambda R \cos \psi}{2\pi p B \sin(\psi + \beta)} \right| \cdot \sigma_{\phi_{ab}} \quad (8.47)$$

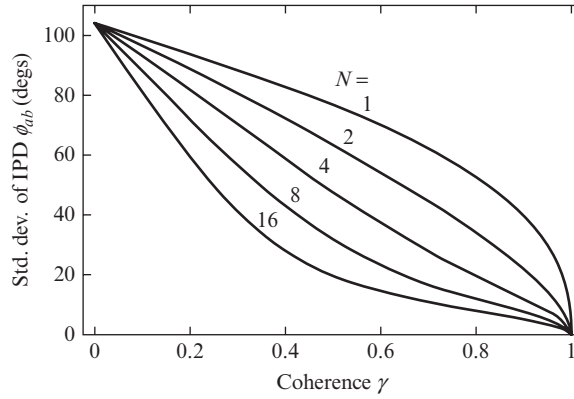
The PDF of the IPD as a function of γ is complicated; it is given in [15,36]. The maximum likelihood estimate of the IPD $\hat{\phi}_{ab}$ was given in equation (8.33); its PDF is also complicated. However, the Cramér-Rao lower bound (CRLB) of the IPD estimate is [17]

$$\sigma_{\hat{\phi}_{ab}}^2 = \frac{(1 - |\gamma|^2)^2}{2N|\gamma|^2} \quad (8.48)$$

where N is the number of interferogram samples available for averaging, often called the number of looks. The CRLB gives a very good estimate of the actual IPD variance for $\gamma \geq 0.5$ and $N \geq 8$. It shows that N -fold averaging reduces the phase variance, and therefore the elevation variance, by the factor N .

Multiple looks can be obtained by dividing the SAR data bandwidth into N subbands, forming reduced-resolution images from each, and averaging the interferograms or by local spatial averaging of a single full-bandwidth interferogram. The latter technique is

FIGURE 8-25 ■ Standard deviation of the interferometric phase estimate as a function of coherence and the number of looks averaged. (From Bamler and Hartl [36]. With permission.)



most commonly used, typically with a 3×3 , 5×5 , or 7×7 window. With either method, the interferogram will be of lower resolution than the full-resolution InSAR image.

Figure 8-25 shows the standard deviation in degrees of the IPD estimate as a function of the coherence and the number of independent looks averaged. For reasonably high coherence values of 0.6 or more, the standard deviation of $\hat{\phi}_{ab}$ decreases particularly quickly as the number of looks is increased from 1 to 4; the rate of improvement slows after that. This suggests that InSAR systems should be designed to average at least four looks in estimating the IPD [17].

As an example, for an SNR of 10 dB equation (8.43) gives a value of $\gamma_{noise} = 0.9$. From Figure 8-25, this will result in an IPD standard deviation of about 40° for one look and about 12° for four looks.

Estimation of the magnitude of the coherence is also of interest, both as a measure of local interferogram quality and in scatterer classification. The maximum likelihood estimator for $|\gamma|$ and its CRLB are [36]

$$|\hat{\gamma}| = \frac{\left| \sum_{n=0}^{N-1} a[n] b^*[n] \right|}{\sqrt{\left(\sum_{n=0}^{N-1} |a[n]|^2 \right) \left(\sum_{n=0}^{N-1} |b[n]|^2 \right)}} \quad (8.49)$$

$$\sigma_{|\hat{\gamma}|}^2 = \frac{(1 - |\gamma|^2)^2}{2N} \quad (8.50)$$

Again, the coherence CRLB is a good estimate of the actual variance for $\gamma \geq 0.5$ and $N \geq 8$.

8.6.1.6 Coherence and Interferometric System Design

It is apparent from equation (8.48) that it is important to design an InSAR system to maximize the coherence, at least on relatively stable scattering surfaces. Low coherence results in large phase errors and thus large elevation errors while also increasing errors in the magnitude of the coherence function, which degrades scatterer classification, change detection, and similar functions. While part of the coherence loss is due to the physical nature of the scatterers and thus beyond designer control, other factors can be influenced by the system design.

Equations (8.45) and (8.46), for example, show that the baseline decorrelation contribution to reduced coherence can be minimized by preferring shallower grazing angles, longer ranges, finer resolution, single transmitter operation ($p = 1$), and a shorter baseline B (and thus shorter B_{\perp}). Equations (8.16) and (8.28) show that a short baseline is also good for reducing elevation ambiguities and the effects of IPD measurement errors but that a large baseline is preferred for increasing the sensitivity of elevation measurements. Temporal decorrelation is minimized by using one-pass or formation flying systems in preference to repeat-pass systems. Coherence loss due to noise is reduced by increasing the signal-to-noise ratio. Of course, all of these choices represent engineering trade-offs.

8.6.2 Systematic Errors

Many InSAR references present a model for the elevation error as a function of various geometric and system parameters. Good introductions are given in [16,17]. Here we give only the effects of sensor baseline and tilt errors as examples, with a brief mention of several additional systematic error sources.

8.6.2.1 Sensor Geometry Errors

Baseline errors include errors in the knowledge of the length B and tilt β of the InSAR baseline. This is primarily an issue for repeat-pass systems. One-pass systems may suffer errors in the knowledge of tilt, but the baseline length is generally accurately known. The sensitivity of elevation to baseline length and orientation can be established by using equations (8.8) and (8.4):

$$\begin{aligned} \frac{dh}{dB} &= \frac{dh}{d(\delta R)} \cdot \frac{d(\delta R)}{dB} \\ &\approx \frac{R \cos \psi}{B \sin(\psi + \beta)} \cdot \cos(\psi + \beta) = \frac{R}{B} \frac{\cos \psi}{\tan(\psi + \beta)} \end{aligned} \quad (8.51)$$

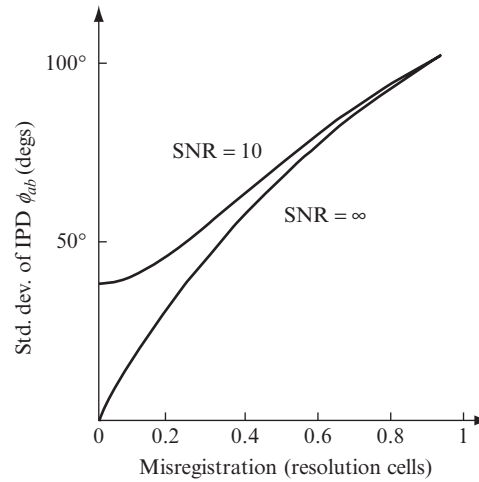
$$\begin{aligned} \frac{dh}{d\beta} &= \frac{dh}{d(\delta R)} \cdot \frac{d(\delta R)}{d\beta} \\ &\approx \frac{R \cos \psi}{B \sin(\psi + \beta)} \cdot \{-B \sin(\psi + \beta)\} = -R \cos \psi \end{aligned} \quad (8.52)$$

Unlike statistical errors, elevation errors due to baseline uncertainties are systematic, biasing each pixel deterministically, though by differing amounts since the grazing angle ψ varies with range.

8.6.2.2 Other Systematic Error Sources

There are numerous additional systematic error sources. Just and Bamler [15] give estimates of the effect of various processing errors such as registration misalignment in range or cross-range, defocusing in range or cross-range, and residual quadratic phase errors on the IPD and thus the elevation estimate. Figure 8-26 gives the effect of misregistration on the standard deviation of ϕ_{ab} . Limiting the IPD standard deviation to 42° , similar to the noise contribution for an SNR of 10 dB, requires that the misregistration be limited to 0.1 resolution cells. Multilook averaging will relax this requirement, while higher quality DEM requirements will tighten it.

FIGURE 8-26 ■
Standard deviation
of the interferometric
phase estimate as a
function of
misregistration.
(From Just and
Bamler [15]. With
permission.)



In some cases, the actual change in IPD from one pixel to the next will exceed 2π radians because the terrain changes elevation by more than h_{amb} meters. An obvious example is at the side of a tall building. In some scenes, it is possible that a region of the interferogram will be completely enclosed by such aliased IPD values: the tall building will have IPD jumps of more than 2π on all four sides. In this event, the phase unwrapping algorithm has no way to determine the correct number of cycles to add to the phase to determine the elevation of the building top relative to the terrain around it. One solution to this problem is multibaseline interferometry, where one baseline is small enough to give a value of h_{amb} sufficiently large to unambiguously determine the elevation of any features in the terrain to be imaged.

Many SAR image formation algorithms such as the range-Doppler or polar format algorithms approximate the spherical wavefronts of the radar signal as planar across the imaged scene, while others such as range migration or backprojection algorithms account for wavefront curvature in the focusing process. Even when an imaging algorithm that accounts for curvature is used, the differential wavefront curvature due to slightly different vantage points contributes a term to the IPD that is not accounted for in many InSAR systems. Wavefront curvature is significant only in systems where the ratio of nominal range to scene extent is relatively small and thus is mainly a problem for short range airborne InSAR. It is corrected by computing the differential path length difference from the platform to each pixel and removing the corresponding phase difference in a manner similar to flat earth phase correction.

8.7 | SOME NOTABLE INSAR SYSTEMS

InSAR has been demonstrated in a variety of spaceborne and airborne systems. While this list is in no way complete, a brief description of a few well-known systems, most already mentioned, follows. Table 8-3 lists approximate values for some of the major parameters of each of these systems. These parameters are considered approximate because of limited information regarding the definitions or means of measurement and inconsistent units and usage among various sources in readily available literature. The references and websites cited provide more information about each system and are current as of this writing (summer 2011).

TABLE 8-3 ■ Approximate Parameters of Some Representative InSAR Systems

	Airborne Systems				Spaceborne Systems						
	CCRS C-SAR	InSARE/ STAR-3i	GeoSAR	RTV	RADARSAT 1	SRTM	ERS-1/2 Tandem	ENVISAT ASAR	ALOS PALSAR	RADARSAT 2	TerraSAR-X Tandem-X
<i>RF</i>	5.3 GHz	9.6 GHz	353 MHz, 10 GHz	16.7 GHz	5.3 GHz	5.3 GHz, 9.6 GHz	5.3 GHz	5.3 GHz	1.27 GHz	5.405 GHz	9.65 GHz
<i>Altitude</i>	6.4 km	6–12 km	10–12 km	5.5–7 km	793–821 km	233 km	785 km	780–820 km	692 km	798 km	514 km
<i>One- or two-pass</i>	1 pass	1 pass	1 pass	1 pass	2 pass	1 pass	1 pass formation flying	2 pass	2 pass	2 pass	1 pass formation flying
<i>Cross-range resolution</i>	6–10 m	1.25 m	1–2 m	0.45–1.1 m	8–50 m	30 m	25 m	6–30 m	7–100 m	3–100 m	1–16 m
<i>Relative vertical accuracy</i>	1.5–5 m RMS	0.5–1.25 m RMS	X: 0.5–1.2 m UHF/P: 1–3 m	1 m LE90	15–50 m RMS	X Band: 6 m (90%)	11–14 m RMS		7–100 m	2–11 m (1σ)	2–4 m
<i>Baseline length</i>	2.8 m	0.92 m	X: 2.6 m UHF/P: 20 m	0.33 m	30–1100 m	60 m	50–500 m	10–500 m	70–850 m	30–1100 m	300–500 m
<i>Baseline orientation from horizontal</i>	59°	Horizontal (0°)	Horizontal (0°)	Vertical (90°)	40–70°	30–75°	67°	45–75°	50–70°	42–49°	variable
<i>Polarizations</i>	C: HH+HV, VV+VH X: HH	HH	X: VV UHF/P: HH+HV, VV+VH	VV	HH	HH, VV	VV	HH, VV, HH+HV, VV+VH	HH, VV, HV, VH, HH+HV, VV+VH, HH+VV+V H+HV	HH, VV, HV, VH, HH+HV, VV+VH, HH+VV+V VH+HV	HH, VV, HV, VH
<i>First year of operation</i>	1991	≈ 1995	2003 (commercial operation)	2001	1995	2000	1995	2002	2006	2007	2007

8.7.1 Spaceborne Systems

The Shuttle Radar Topography Mission (SRTM) The InSAR system with the greatest public awareness is certainly the space shuttle-based SRTM system [9,10]. The SRTM refers to the specific 11-day mission flown by the space shuttle Endeavour in February 2000. The radar was a dual C and X band InSAR. One aperture for each band was located in the shuttle cargo bay, while the other was at the end of a 60 m rigid mast extending from the bay. The SRTM mission acquired the data needed to map approximately 80% of the Earth's surface to DTED level 2 specifications in 11 days. Extensive information on the SRTM mission, including imagery and educational materials, can be found online (<http://www2.jpl.nasa.gov/srtm/>).

Unedited C band data were completed and released to NGA in January 2003, which in turn edited and verified the data and formatted it into compliance with DTED standards. This task was finished in September 2004. SRTM-derived C band DTED level 1 and level 2 data are now publicly available through the U.S. Geological Survey's EROS Data Center (http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/Elevation_Products). X band data were processed by the German Aerospace Center (DLR) to DTED level 2 specifications and are available from DLR; more information can be found online (http://www.dlr.de/srtm/level1/products_en.htm).

RADARSAT 1 and 2: The Canadian Centre for Remote Sensing (CCRS) launched the RADARSAT 1 Earth observation satellite SAR in 1995 [59]. RADARSAT 1 is a C band, single polarization (HH) system. Its primary mission is to monitor environmental change and support resource sustainability. RADARSAT is also a major source of commercially available satellite SAR imagery. Though not initially designed for InSAR usage, the system is now routinely used in a repeat-pass mode for InSAR. Orbital and operational considerations result in time between SAR image pairs on the order of days to months. RADARSAT 2, launched in December 2007, extended the data products produced by RADARSAT 1 by adding a capability for full polarimetric scattering matrix (PSM) collection, improving the best-case resolution from 8 to 3 m, and introducing additional imaging modes. A RADARSAT Constellation program is planned to expand to three satellites by 2015. Information on RADARSAT 1 and 2 and RADARSAT Constellation is available online (<http://www.asc-csa.gc.ca/eng/satellites/>).

ERS-1 and ERS-2: The European Space Agency (ESA) developed the European Remote Sensing (ERS) 1 and 2 satellite SAR systems launched in 1991 and 1995, respectively [60–62]. Of special interest for InSAR is the ERS 1/2 tandem mission, in which the satellites fly in the same orbital plane and mean altitude, with their orbits phased to have the ERS-2 ground track follow that of ERS-1 with a 24-hour time lag [60]. This provides the global interferometric coverage of a spaceborne InSAR with a much shorter temporal baseline than can be supported by RADARSAT, greatly reducing coherence loss due to temporal decorrelation. Information on ERS 1 and 2 is available online (<http://earth.esa.int/ers/>).

ENVISAT: ERS-1/2 were succeeded by ESA's ENVISAT satellite, which carries 10 Earth-monitoring instruments, among them the Advanced SAR (ASAR) [63]. Information on the ENVISAT ASAR is available online (<http://envisat.esa.int/instruments/asar/>).

ALOS PALSAR: PALSAR is a full-polarimetric L band SAR on the Japanese JERS-1 satellite. The system was developed by the Earth Remote Sensing Data Analysis Center

(ERSDAC), which distributes data products from PALSAR and other ERSDAC systems. Information on the PALSAR project is available online (<http://www.palsar.ersdac.or.jp/e/>).

TerraSAR-X. Launched in 2007, TerraSAR-X is an X-band full-polarization satellite operating at a relatively low orbit of 514 km [64]. The system was developed and launched by a partnership consisting of DLR and EADS Astrium GmbH. Data exploitation rights are held by Infoterra GmbH. In 2010, a second satellite TanDEM-X was launched, enabling the generation of high-resolution DEMs on a global scale. Information on TerraSAR-X and TanDEM-X is available online (http://www.dlr.de/eo/en/desktopdefault.aspx/tabid-5725/9296_read-15979/ and <http://www.infoterra.de/terrasar-x-satellite>). Extensive information about the satellite, sensor, and ground system is available in the special issue on TerraSAR-X of the *IEEE Transactions on Geoscience and Remote Sensing* [65].

8.7.2 Airborne Systems

CCRS C/X-SAR: The CCRS has operated an airborne C band SAR since 1986; X band was added in 1988, and a second antenna to support one-pass InSAR at C band was added in 1991 [66,67]. The system is mounted on a Convair 580 aircraft. It has been used by the remote sensing research and development community, resource managers, and the exploration, maritime, and mapping industries and it supported initial design and marketing for the RADARSAT space-based radar. The CCRS system has been particularly heavily used for along-track interferometry research, especially as applied to mapping ocean currents and glacial movement.

InSARE/STAR-3i: Two InSAR elevation (InSARE) systems were developed under the sponsorship of the U.S. Advanced Research Projects Agency (ARPA, now DARPA) in 1992–1993 by Norden Systems, Inc. (now part of Northrop Grumman Corporation) and the Environmental Research Institute of Michigan (ERIM, now part of General Dynamics Corporation) [68,69]. The ERIM InSARE discussed here is now operated by Intermap Technologies, Ltd. and is called the STAR-3i system. It is an X band system emphasizing relative rapid generation of digital elevation data for such purposes as site surveys and monitoring for construction and environmental purposes, obtaining elevation data in areas where changes have occurred, tactical military applications, and others.

GeoSAR: GeoSAR is a dual-frequency P (low ultra high frequency) and X band InSAR for environmental management and geological, seismic, and environmental hazard identification and monitoring [70,71]. Developed by the U.S. Jet Propulsion Laboratory working with Calgis, Inc. and the California Department of Conservation, the system is intended to provide both top-of-canopy DSMs with the X band InSAR and bare Earth DEMs using the P band InSAR. Similar to the SRTM, a laser ranging system is used to aid in baseline length and orientation estimation. The system is now operated by Fugro, Inc. Additional information on the GeoSAR system is available online (<http://www.geosar.com/>).

The Rapid Terrain Visualization System: The RTV system was developed by Sandia National Laboratories for the U.S. Army with the purpose of “rapid generation of digital topographic data to support emerging crisis or contingencies” [7]. The RTV is a K_u band system flown on a deHavilland DHC-7 aircraft. The system has at least two unique aspects. The first is the use of an elevation monopulse antenna for one of the apertures to enable

multibaseline InSAR processing, eliminating the need for explicit phase unwrapping. The second is real-time onboard generation of mosaiced InSAR data products at peak area mapping rates of 10 km²/minute (DTED level 3) or 3.5 km²/minute (DTED level 4). Additional information is available online (<http://www.sandia.gov/radar/rtv.html>).

8.8 | OTHER COHERENT EXPLOITATION TECHNIQUES

InSAR is a form of coherent exploitation of SAR imagery. This term refers to the use of multiple complex SAR images, collected from different vantage points, times, or both to form new data products.¹⁵ The distance between vantage points is called the spatial baseline, while the difference in vantage times is called the temporal baseline. Spatial baselines are further subdivided into along-track (collinear with the radar velocity vector) and cross-track (orthogonal to the velocity vector) components.

The InSAR technique for measuring elevation variations discussed so far in this chapter uses images displaced along a cross-track spatial baseline and is sometimes referred to as cross-track interferometry (CTI). In this section, we briefly introduce three other coherent SAR exploitation techniques. Terrain motion mapping and change detection both employ images taken from the same vantage point but at significantly different times and thus use a temporal baseline but no spatial baseline. Change detection generally implies time baselines of minutes to perhaps days, while terrain motion mapping typically implies time periods of days to years. Because the data are obtained by repeat-pass operation along the same flight path, these are both examples of repeat-pass interferometry.

The third exploitation method introduced here uses images generated from apertures separated by an along-track spatial baseline and is referred to as along-track interferometry. Due to the platform motion, ATI can also be viewed as short (less than 1 second, typically) temporal baseline processing. It is used for measuring time-varying elements in the SAR scene, such as mapping of ocean wave fields, glacier movements, and detection of moving surface targets. A significant extension of this concept enables the combination of SAR imaging with interference cancellation using space-time adaptive processing techniques. A good introduction and comparison of spatial, temporal, and mixed baseline radar applications is given in [51].

8.8.1 Terrain Motion Mapping

InSAR presumes that there is no change in the imaged scene between the two image data collections, so phase differences are due only to elevation variations viewed from slightly different aspect angles. This assumption is clearly true in one-pass systems but may not be in repeat-pass systems due to temporal decorrelation. This problem can also be an opportunity: methods similar to InSAR can be used to detect changes in terrain elevation over significant time periods.

Terrain motion mapping examines the change in phase due to a change in scatterer elevation at a fixed location on the ground between two different times [20]. As with

¹⁵The effective vantage point and time for a SAR image are, respectively, the position and time of the radar platform at the center of the synthetic aperture from which the image is formed.

InSAR static terrain mapping, we again assume that the reflectivity $\rho(x, y)$ of each pixel does not change between images. Two images $a(x, y)$ and $b(x, y)$ of a scene are collected from identical vantage points (no spatial baseline) and coregistered. If the terrain elevation changes between the two acquisition times but the reflectivity does not, then the only term in the pixel phase of equation (8.10) that will change between images a and b is the slant range R . Denoting the change in range along the LOS as R_{ab} , the phase difference will then satisfy

$$\phi_{ab} = -\frac{4\pi}{\lambda}(R_a - R_b) = -\frac{4\pi}{\lambda}\delta R_{ab} \quad (8.53)$$

For a scatterer at a given ground range, and assuming planar wavefronts for simplicity, a change in elevation of δh will result in a change in LOS range of $\delta h \cdot \sin \psi$. Consequently, the elevation change can be estimated as

$$\widehat{\delta h}(x, y) \approx \frac{\lambda}{4\pi \sin \psi} \phi_{ab}(x, y) \quad (8.54)$$

This equation assumes that the terrain motion between passes is in the vertical dimension only. In fact, in many cases, such as earthquakes or glacial flows, the motion is primarily horizontal. Any change in slant range between a scatterer and the radar will result in a detectable change in the IPD, and suitable generalizations of equation (8.54) can be developed for horizontal motion of the terrain.

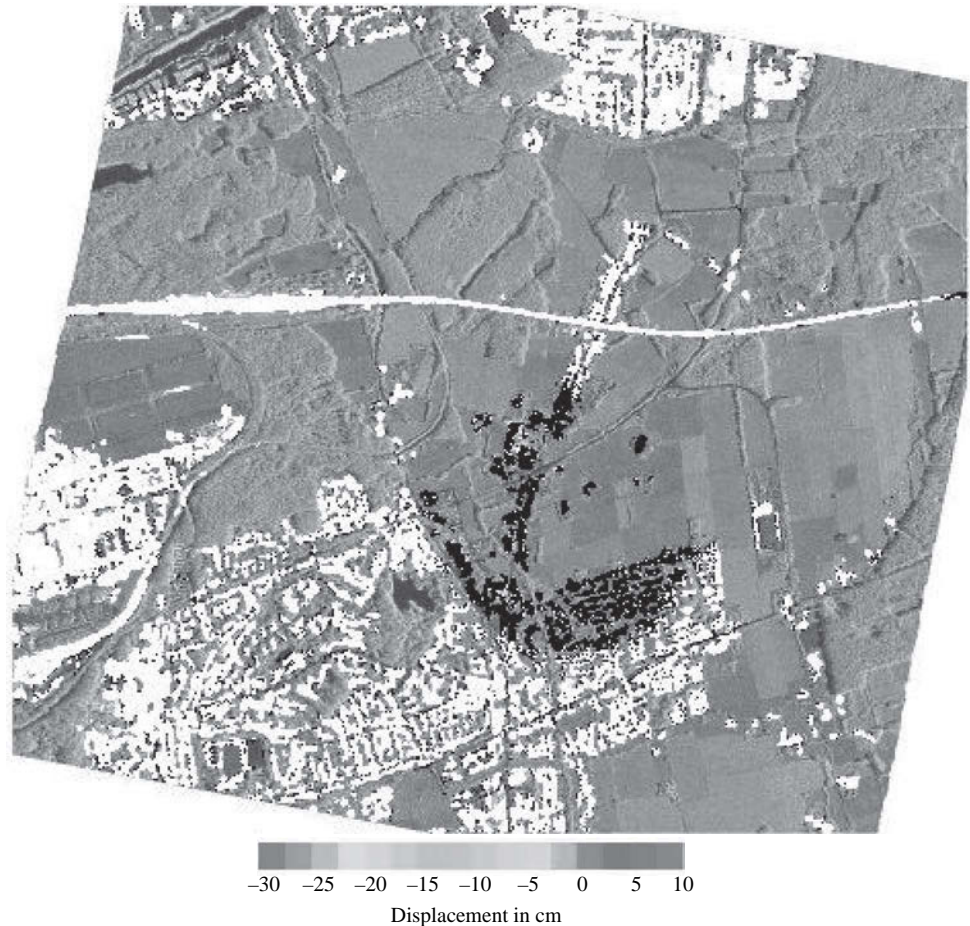
Clearly, terrain motion mapping requires repeat-pass operation. The time interval could be on the order of days or weeks to study the effects of such phenomena as earthquakes or volcanic explosions, or it could be years to study phenomena such as glacier movement or ground subsidence. The processing operations are essentially the same as discussed earlier.

Figure 8-27 is an example of using terrain motion mapping to monitor land subsidence. The map is of the Las Vegas, Nevada, area and is based on data collected over nearly four years. Figure 8-27a is the InSAR terrain motion map; outlines have been added to indicate the areas of greatest subsidence. Notice the sensitivity of the technique: the



FIGURE 8-27 ■ InSAR-based terrain motion map showing land subsidence in Las Vegas, Nevada, over a nearly four-year period. (a) InSAR-based terrain motion map. Subsidence is significant primarily in the outlined regions, with the subsidence being greatest in the darker region in the upper left quadrant. (b) Three-dimensional visualization of subsidence effects. (Courtesy of the Radar Interferometry Group, Stanford University.)

FIGURE 8-28 ■
Terrain motion mapping using persistent scatterer interferometry. Here PSI is used to map subsidence over approximately 10 months due to underground mining in the Ruhr area of Germany. White pixels are persistent scatterers with elevation changes of less than 10 cm. Black pixels are persistent scatterers with more than about 10 cm subsidence. See the online version for a more detailed color scale. (From Wegmüller et al. [73]. With permission.)



subsidence is less than 10 centimeters over four years. The subsidence in this case is due to inelastic compaction of the aquifer [72]. Figure 8-27b is a rather dramatic three-dimensional visualization generated from this data.

Figure 8-28 is a newer example. Persistent scatterer interferometry was used to detect land subsidence due to underground mining using collections over a period of about 10 months [73]. Black pixels show persistent scatterers that subsided about 10 cm or more. The online color version of this figure provides additional detail.

Ideally, exactly the same flight path would be followed on the two passes, so the baseline between the two images is zero. In practice, this is very difficult, and a small nonzero baseline will be reflected in the data. This means the IPD will have components due to both the temporal change in elevation and the static elevation profile. One approach to removing the static component is to use an existing DEM to estimate it and then subtract its contribution to the IPD, a method known as differential InSAR [20,23].

8.8.2 Coherent Change Detection

Another coherent exploitation technique with increasing applications is change detection. Like terrain motion mapping, change detection is a repeat-pass application that compares two or more images taken from the same trajectory at different times. The time intervals are

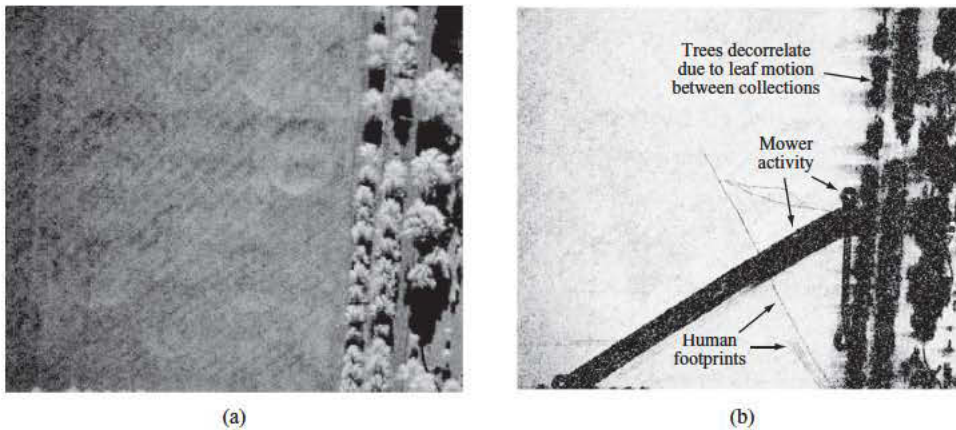


FIGURE 8-29 ■ InSAR-based coherent change detection. (a) One of a pair of SAR images of a field bordered by trees. (b) CCD change map showing mower activity and footprints of pedestrians. The trees decorrelate due to leaf motion. (Courtesy of Sandia National Laboratories. With permission.)

typically shorter, from a few minutes apart to many hours or days apart. Change detection assumes that the general terrain elevation profile is unchanged, but the reflectivity function $\rho(x, y)$ does change. Reflectivity changes could occur due to disturbance of the ground by persons or vehicles but also due to wind blowing tree leaves and other natural phenomena. They could also represent the movement of vehicles into or out of a scene.

If there is no change in the reflectivity of a given pixel between passes, computing the magnitude of the coherence $|\gamma(x, y)|$ of the two measurements $\rho_1(x, y)$ and $\rho_2(x, y)$ should produce a value approximately equal to 1.0. If the reflectivity has changed, a lesser value of coherence should result. A good estimator for $|\gamma|$ was given in equation (8.49). Typical averaging windows range in size from 3×3 to 9×9 so that N ranges from about 9 to 81.

A coherent change detection (CCD) map can provide a very sensitive indicator of activity in an observed area and is also effective at removing background clutter prior to classification of features in SAR imagery. Figure 8-29 is an example of the first application. Figure 8-29a shows one of a pair of SAR images of a field bordered by a stand of trees; the second image would appear virtually identical to the eye. The data for the two images were collected on flight passes separated by approximately 20 minutes. Figure 8-29b is the CCD change map formed from the image pair. Light-colored pixels represent values of $|\gamma|$ near 1.0, and dark pixels represent values near 0.0. The map clearly shows a broad diagonal dark streak and another vertical, narrower streak where mowers had cut the field between the two SAR passes, changing its reflectivity. Also visible are some narrow trails where pedestrians walked in the scene, disturbing the field surface. Note that the trees also decorrelated (low coherence) between passes. This is due to wind blowing the individual leaves that comprise the composite response of each pixel, effectively randomizing the pixel reflectivity phases between passes.

8.8.3 Along-Track Interferometry

ATI compares the pixel-by-pixel phases of two SAR images collected from the same vantage points but separated by a short time interval [51]. The short time interval is achieved by using two receive apertures on the same platform, displaced along the velocity vector

instead of orthogonal to it as in InSAR. If the apertures are separated by Δx m along the platform track and the platform velocity is v m/s, then the time lag between apertures will be $\Delta x/v$ seconds. For a typical airborne SAR with $v = 200$ m/s and along-track baselines of 2 to 20 m, the time lag is 0.01 to 0.1 seconds. Thus, while the terrain motion mapping and change detection techniques previously discussed also involve comparing images for the same vantage point but different times, in ATI the time lags are much shorter, fractions of a second instead of hours, weeks, or years. Consequently, ATI is useful for mapping active dynamic processes on the terrain surface and for detecting moving targets in the terrain, situations involving relatively rapid changes in the scene.

Consider a scatterer moving with a radial velocity component v_r along the sidelooking LOS between the platform track and the resolution cell of interest, with positive v_r corresponding to increasing range, that is, movement away from the radar. In the time lag of $\Delta x/v$ between the two synthetic apertures, the scatterer will be displaced by $v_r \cdot \Delta x/v$ m. With $a(x, y)$ representing the image from the leading aperture and $b(x, y)$ that from the trailing aperture, $R_b = R_a + v_r \cdot \Delta x/v$. From equation (8.53) the IPD for that resolution cell becomes

$$\phi_{ab} = -\frac{2\pi p}{\lambda}(R_a - R_b) = +\frac{2\pi p v_r \Delta x}{\lambda v} \quad (8.55)$$

where we have again generalized to allow for either one- or two-transmitter operation. The former case occurs when there are two independent simultaneous receive apertures displaced along-track on the platform but only one transmits. The latter case applies when there is only one transmitter and receiver. The radial velocity can be estimated as

$$v_r = \frac{\lambda v}{2\pi p \Delta x} \phi_{ab} \quad (8.56)$$

The precision with which this IPD can be measured is determined by the coherence. For the short baselines and time lags generally used in ATI, the principal concern is additive noise. At an SNR of 20 dB, the coherence $\gamma_{noise} = 0.99$ and, from Figure 8-25, the standard deviation of the IPD for a single look will be about 6° or 0.1 radians; for SNR = 10 dB, these increase to about 40° (0.7 radians). Using equation (8.56) with the platform parameters above and $p = 1$, the corresponding precision in the estimate of v_r is about 0.08 m/s at 20 dB and 0.56 m/s at 10 dB. Averaging N multiple looks will improve these values by approximately a factor of N for small (<10) N . Note also that larger baselines (bigger Δx) decrease the standard deviation of v_r , allowing detection of slower velocities, while shorter baselines are more suited to larger velocities.

The TanDEM-X mission, with its use of two formation-flying satellite and two subarray outputs on each satellite, has four along-track apertures as shown in Figure 8-30. The two on-satellite apertures are 2.4 m in length, while the intersatellite along track spacing can be adjusted from nearly zero to several kilometers [56]. This allows the creation of both short and long temporal baselines to support detection of a wide range of velocities and ambiguous phase resolution for high-velocity scatterers.

As described, ATI measures the along-LOS velocity component of the scatterers. This can be projected into ground range and vertical components using knowledge of the grazing angle ψ . One way to obtain along-track velocity components is to use a squinted (non-sidelooking) beam.

One of the major applications of ATI is the measurement of ocean and tidal currents. Figure 8-31a is an example of mapping of ocean current fields in a 2×2 km region around

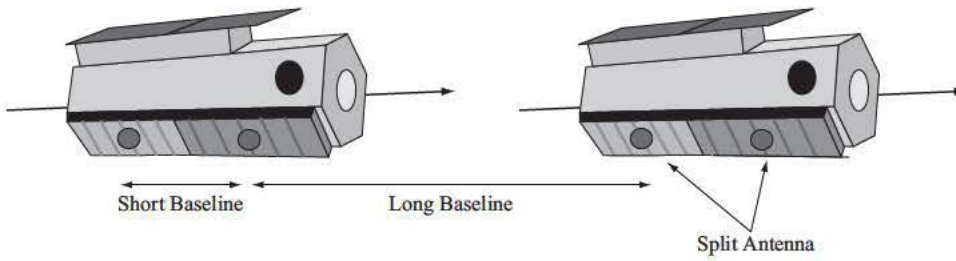


FIGURE 8-30 ■ Along-track baselines available with the TanDEM-X formation flying mission. (From Krieger et al. [56]. With permission.)

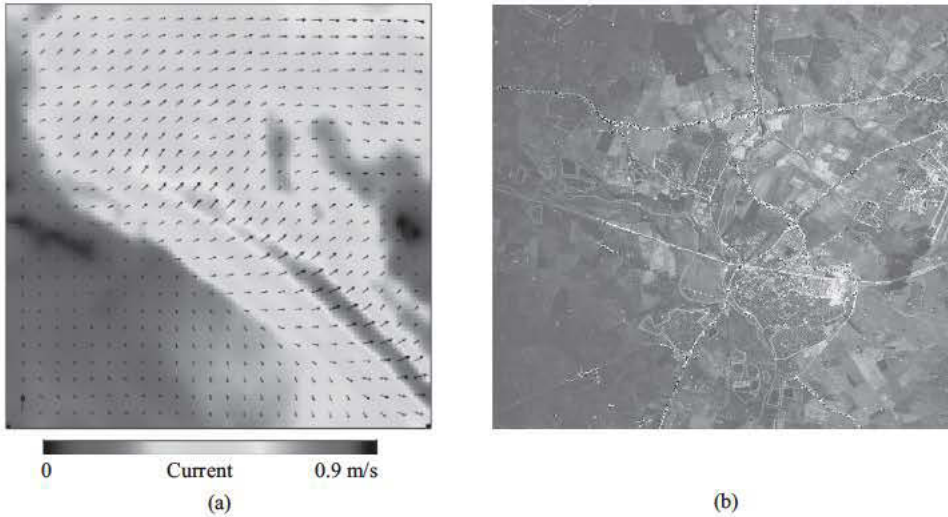


FIGURE 8-31 ■ Two example applications of along-track interferometry. (a) Mapping of the ocean current field around the Heligoland reef by an airborne ATI system. (b) Detection and geolocation of ground moving targets from the Joint STARS surveillance aircraft. Black or white pixels indicate moving target detections. See the online color version for additional detail. (Image a from Romeiser and Hirsch [74]. With permission. Image (b) is U.S. government work, not protected by copyright.)

the Heligoland reef by an airborne ATI system [74]. If two or more different temporal baselines can be established, it is possible to estimate the decorrelation time of the ocean or another moving surface. The coherence (correlation coefficient) of a cell is measured by computing the IPD of image pairs at different temporal lags; two lags are typical. It is expected that the coherence will be less at the longer time lag than at the shorter. A functional form is assumed for the coherence behavior with time, for example, a decaying exponential:

$$\gamma(B_t) = \gamma_0 e^{-B_t/B_{tc}} \quad (8.57)$$

Here B_t is the temporal baseline (time lag), and γ_0 is the coherence at zero time lag, which is determined primarily by noise so that $\gamma_0 \approx \gamma_{noise}$. B_{tc} is the time constant for the decorrelation and is the parameter of interest. Given two coherence measurements B_{t1} and B_{t2} , B_{tc} can be estimated as [51]

$$B_{tc} = \frac{B_{t2} - B_{t1}}{\ln[\gamma(t_1)/\gamma(t_2)]} \quad (8.58)$$

An early but very successful use of ATI for moving target detection is the U.S. Joint STARS airborne surveillance radar system. First pressed into service in the 1991 Gulf

War, the system can image terrain and detect and geolocate ground moving vehicle targets from substantial standoff ranges. Figure 8-31b shows an example Joint STARS GMTI data product [75]. Each of the black or white dots represents a ground moving target detected by the system, geolocated, and overlaid onto an image of the terrain. The online color version provides additional detail.

8.9 | SUMMARY

This chapter has shown how pairs of complex SAR images can be combined to generate several useful remote sensing products. The focus has been primarily on digital elevation maps, but it has also introduced ways to measure temporal changes in the profile or reflectivity of a scene on both short and long time scales using along-track interferometry, coherent change detection, and temporal motion mapping. All of these techniques rely on measurements of interferometric phase differences between images, differing in whether and how much those images are separated along spatial and temporal baselines. Because they all rely on IPD as the fundamental measured quantity, they share many common signal processing steps: subpixel image coregistration; two-dimensional phase unwrapping or its equivalent; filtering and multilook averaging for maximizing coherence; and orthorectification and geocoding to provide useful final data products.

These mission concepts and processing techniques continue to be active research areas: InSAR is expanding into true 3-D imaging using Fourier, tomographic, and other methods, while ATI is expanding into combined GMTI and SAR. More advanced SAR sensors and the advent of formation and constellation sensor systems continue to expand the menu of measurement configurations. These developments mark progress toward a capability for timely, high-quality earth resources data acquisition as well as military and security surveillance, both on a global scale.

8.10 | FURTHER READING

An excellent first introduction to the concepts and issues in InSAR is given by Madsen and Zebker in [51]. Detailed tutorial developments of InSAR with an airborne radar perspective are given in the spotlight SAR textbooks by Jakowatz et al. [20] and Carrara et al. [14]. An analysis from a spaceborne radar perspective is given in the book by Franceschetti and Lanari [33]. The tutorial paper by Rosen et al. [24] also emphasizes spaceborne systems and provides a good overview of space-based InSAR applications as well as an extensive bibliography. Bamler and Hartl [36] is another excellent tutorial paper, again with a spaceborne emphasis. Additional tutorial sources are [76,77]. Early attempts at interferometric radar are described in [78-80]. The first descriptions of the use of coherent imagery for InSAR were reported in [81,82]. The first InSAR-related patent application was apparently that of D. Richman, then at United Technologies Corp. [83]. The application was filed in 1971 but was placed under a secrecy order and was not granted until 1982 [51].

While not tutorial in nature, a comprehensive set of papers describing virtually all aspects of a modern remote sensing satellite system is the special issue on TerraSAR-X of the *IEEE Transactions on Geoscience and Remote Sensing* [65]. The papers in this issue describe in detail the mission, satellite design, ground system design, calibration procedures, and early application results.

8.11 | REFERENCES

- [1] Brenner, A.R. and Roessing, L., “Radar Imaging of Urban Areas by Means of Very High-Resolution SAR and Interferometric SAR”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2971–2982, October 2008.
- [2] El-Sheimy, N., Valeo, C., and Habib, A., *Digital Terrain Modeling: Acquisition, Manipulation, and Applications*, Artech House, Norwood, MA, 2005.
- [3] U.S. Geological Survey. “US GeoData Digital Elevation Models Fact Sheet.” Available at <http://erg.usgs.gov/isb/pubs/factsheets/fs04000.html>.
- [4] Melvin, W.L., Showman, G.A., and Guerci, J.R., “A Knowledge-Aided GMTI Detection Architecture,” in *Proceedings of the IEEE Radar Conference*, Philadelphia, PA, April 26–29, 2004.
- [5] U. S. National Geospatial-Intelligence Agency, “Digital Terrain Elevation Data (DTED)”, available at <https://www1.nga.mil/ProductsServices/TopographicalTerrestrial/DigitalTerrainElevationData/Pages/default.aspx>.
- [6] U.S. National Geospatial-Intelligence Agency, “Performance Specification, Digital Terrain Elevation Data (DTED),” MIL-PRF-89020B, May 23, 2000.
- [7] Burns, B.L., Eichel, P.H., Hensley, W.H., and Kim, T.J., “InSAR for the Rapid Terrain Visualization Demonstration,” in *Conference Record of Asilomar Conference on Signals, Systems, and Computers*, vol. 1, pp. 8–15, Pacific Grove, CA, October 2000.
- [8] Roth, M.W., “High-Resolution Interferometric Synthetic Aperture Radar for Discoverer II,” *Johns Hopkins APL Technical Digest*, vol. 20, no. 3, pp. 297–304, 1999.
- [9] Rabus, B., et al., “The Shuttle Radar Topography Mission – A New Class of Digital Elevation Models Acquired by Spaceborne Radar,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 57, pp. 241–262, 2003.
- [10] Shuttle Radar Topography Mission (SRTM), Jet Propulsion Laboratory, National Aeronautics and Space Administration. Available at <http://www2.jpl.nasa.gov/srtm/>.
- [11] National Geospatial-Intelligence Agency, “High Resolution Elevation Product Specification,” specification NGA.IP.0002.1.0. Available at <http://www.gwg.nga.mil/ntb/baseline/docs/HRE.spec/index.html>.
- [12] Richards, M.A., Scheer, J.A., and Holm, W.A., Eds., *Principles of Modern Radar: Basic Principles*, SciTech Publishing, Raleigh, NC, 2010.
- [13] Mercer, J. B., “SAR Technologies for Topographic Mapping,” *Photogrammetric Week 1995*, Mercer, J.B., “SAR Technologies for Topographic Mapping,” pp. 117–126 in Institute for Photogrammetry, *Photogrammetric Week 1995*, Ed. D. Fritsch and D. Hobbie, Stuttgart, Germany.
- [14] Carrara, W.G., Goodman, R.S., and Majewski, R.M., *Spotlight Synthetic Aperture Radar*, Artech House, Norwood, MA, 1995.
- [15] Just, D. and Bamler, R., “Phase Statistics of Interferograms with Applications to Synthetic Aperture Radar,” *Applied Optics*, vol. 33, no. 20, pp. 4361–4368, July 1994.
- [16] H. A. Zebker et al., “Accuracy of Topographic Maps Derived from ERS-1 Interferometric Radar”, *IEEE Transactions Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 823–836, July 1994.
- [17] Rodriguez, E. and Martin, J.M., “Theory and Design of Interferometric Synthetic Aperture Radars,” *IEE Proceedings–F*, vol. 139, no. 2, pp. 147–159, April 1992.

- [18] Itoh, K., "Analysis of the Phase Unwrapping Problem," *Applied Optics*, vol. 2, no. 14, p. 2470, July 15, 1982.
- [19] U.S. National Imagery and Mapping Agency, "Department of Defense World Geodetic System 1984," Technical Report TR 8350.2, 3d ed., Amendment 1, January 1, 2000. Available at http://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html.
- [20] Jakowatz, Jr., C.V., et al., *Spotlight Mode Synthetic Aperture Radar*, Kluwer Academic Publishers, Boston, 1996.
- [21] Richards, M.A., *Fundamentals of Radar Signal Processing*, McGraw-Hill, New York, 2005.
- [22] European Space Agency, *InSAR Principles*, Training Manual TM-19, February 2007.
- [23] Richards, J.A., *Remote Sensing with Imaging Radar*, Springer, 2009.
- [24] Rosen, P.A., Hensley, S., Joughin, I.R., Li, F.K., Madsen, S.N., Rodriguez, E., and Goldstein, R.M., "Synthetic Aperture Radar Interferometry," *Proceedings IEEE*, vol. 88, no. 3, pp. 333–381, March 2000.
- [25] Mercer, B., "Combining LIDAR and IfSAR: What Can You Expect?," *Proceedings Photogrammetric Week 2001*, pp. 227–237, Institute for Photogrammetry, University of Stuttgart. Available at <http://www.intermaptechnologies.com> or <http://www.ifp.uni-stuttgart.de/publications/phowo01/phowo01.en.htm>.
- [26] Mercer, B., "DEMs Created from Airborne InSAR—An Update," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, XXXV-B2, pp. 841–848.
- [27] Goldstein, R., "Atmospheric Limitations to Repeat-Track Radar Interferometry," *Geophysical Research Letters*, vol. 22, no. 18, pp. 2517–2520, 1995.
- [28] Faller, N. and Weber, M., "TerraSAR-X and TanDEM-X: Revolution in Spaceborne Radar," in *IEEE International Geoscience and Remote Sensing Symposium*, pp. 4924–4928, 2007.
- [29] Gamba, P. and Houshmand, B., "Digital Surface Models and Building Extraction: A Comparison of InSAR and LIDAR Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 4, pp. 1959–1968, July 2000.
- [30] Leberl, F.W., *Radargrammetric Image Processing*, Artech House, Norwood, MA, 1990.
- [31] Zitová, B. and Flusser, J., "Image Registration Methods: A Survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.
- [32] Bentoutou, Y., et al., "An Automatic Image Registration for Applications in Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2127–2137, September 2005.
- [33] Franceschetti, G. and Lanari, R., *Synthetic Aperture Radar Processing*, CRC Press, Boca Raton, FL, 1999.
- [34] Imel, D.A., "Accuracy of the Residual-Delay Absolute-Phase Algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 1, pp. 322–324, January 1998.
- [35] Scheiber, R. and Moreira, A., "Coregistration of Interferometric SAR Images Using Spectral Diversity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2179–2191, September 2000.
- [36] Bamler, R. and Hartl, P., "Synthetic Aperture Radar Interferometry," *Inverse Problems*, vol. 14, pp. R1–R54, 1998.
- [37] Ghiglia, D.C. and Pritt, M.D., *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*, Wiley, New York, 1998.

- [38] Zebker, H.A. and Lu, Y., “Phase Unwrapping Algorithms for Radar Interferometry: Residue-Cut, Least-Squares, and Synthesis Algorithms,” *Journal of the Optical Society Of America*, vol. 15, no. 3, pp. 586–598, March 1998.
- [39] Constantini, M., “A Novel Phase Unwrapping Method Based on Network Programming,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 813–821, May 1999.
- [40] Gens, R., “Two-Dimensional Phase Unwrapping for Radar Interferometry: Developments and New Challenges,” *International Journal on Remote Sensing*, vol. 24, no. 4, pp. 703–710, 2003.
- [41] Chen, C.W. and Zebker, H.A., “Network Approaches to Two-Dimensional Phase Unwrapping: Intractability and Two New Algorithms,” *Journal of the Optical Society of America*, vol. 17, no. 3, pp. 401–414, March 2000.
- [42] Loffeld, O., et al., “Phase Unwrapping for SAR Interferometry—A Data Fusion Approach by Kalman Filtering,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 1, pp. 47–58, January 2008.
- [43] Martinez-Espla, J.J., Martinez-Marin, T., and Lopez-Sanchez, J.M., “A Particle Filter Approach for InSAR Phase Filtering and Unwrapping,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1197–1211, April 2009.
- [44] Bian, Y. and Mercer, B., “Interferometric SAR Phase Filtering in the Wavelet Domain Using Simultaneous Detection and Estimation,” *IEEE Transactions Geoscience and Remote Sensing*, vol. 49, no. 4, pp. 1396–1416, April 2011.
- [45] Goldstein, R.M., Zebker, H.A., and Werner, C.L., “Satellite Radar Interferometry: Two-Dimensional Phase Unwrapping,” *Radio Science*, vol. 23, no. 4, pp. 3268–3270, 1989.
- [46] Hensley, S., Rosen, P., and Gurrola, E., “The SRTM Topographic Mapping Processor,” in *Proceedings of the Geoscience and Remote Sensing Symposium*, vol. 3, pp. 1168–1170, July 2000.
- [47] Bonifant, Jr., W.W., Richards, M.A., and McClellan, J.H., “Interferometric Elevation Estimation of the Seafloor via Synthetic Aperture Sonar in the Presence of Motion Errors,” *IEE Proceedings—Radar, Sonar, and Navigation*, vol. 147, no. 6, pp. 322–330, December 2000.
- [48] Ghiglia, D.C. and Romero, L.A., “Robust Two-Dimensional Weighted and Unweighted Phase Unwrapping that Uses Fast Transforms and Iterative Methods,” *Journal of the Optical Society of America*, vol. 11, no. 1, pp. 107–117, January 1994.
- [49] Oppenheim, A.V. and Schaffer, R.W., *Discrete-Time Signal Processing*, 3d ed., section 8.8.2, Pearson, Upper Saddle River, NJ, 2010.
- [50] Zhang, K., et al., “Phase Unwrapping for Very Large Interferometric Data Sets,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 12, pp. 4204–4218, December 2010.
- [51] Madsen, S.N. and Zebker, H.A., “Imaging Radar Interferometry,” Chapter 6 in *Principles and Applications of Imaging Radar—Manual of Remote Sensing*, 3d ed., vol. 2, Ed. F. M. Henderson and A. J. Lewis, Wiley, New York, 1998.
- [52] Madsen, S.N., “On Absolute Phase Determination Techniques in SAR Interferometry,” in *Proceedings SPIE Algorithms for Synthetic Aperture Radar Imagery II*, vol. 2487, pp. 393–401, Orlando, FL, April 19–21, 1995.
- [53] Hein, A., *Processing of SAR Data*, Springer, 2004.
- [54] Zebker, H.A. and Villasenor, J., “Decorrelation in Interferometric Radar Echoes,” *IEEE Transactions on Geoscience Remote Sensing*, vol. 30, no. 5, pp. 950–959, September 1992.
- [55] Ferretti, A., Prati, C., and Rocca, F., “Permanent Scatterers in SAR Interferometry,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 1, pp. 8–20, 2001.

- [56] Krieger, G., et al., "Interferometric Synthetic Aperture Radar (SAR) Missions Employing Formation Flying," in *Proceedings of the IEEE*, vol. 98, no. 5, pp. 816–843, May 2010.
- [57] Hanssen, R.F., *Radar Interferometry: Data Interpretation and Error Analysis*, Springer, New York, 2001.
- [58] D'Addio, S. and Ludwig, M., "Modelling and analysis of rain effect on Ka-band single pass InSAR Performance," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 4, pp. IV-913–IV-916, 2009.
- [59] Geudtner, D., et al., "RADARSAT Repeat-Pass SAR Interferometry," in *Proceedings IEEE 1998 International Geoscience and Remote Sensing Symposium*, vol. 3, pp. 1635–1637, July 6–10, 1998.
- [60] Duchossois, G. and Martin, P., "ERS-1 and ERS-2 Tandem Operations," European Space Agency *ESA Bulletin*, no. 83, pp. 54–60, 1995.
- [61] Rufino, G., Moccia, A., and Esposito, S., "DEM Generation by Means of ERS Tandem Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 6, pp. 1905–1912, November 1998.
- [62] Shipping, S., "DEM Generation Using ERS-1/2 Interferometric SAR Data," *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, vol. 2, pp. 788–790, 2000.
- [63] Suchail, J.-L., et al., "The ENVISAT-1 Advanced Synthetic Aperture Radar Instrument," in *Proceedings IEEE Geoscience and Remote Sensing Symposium*, vol. 2, pp. 1441–1443, 1999.
- [64] Werninghaus, R. and Buckreuss, S., "The TerraSAR-X Mission and System Design," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 2, February 2010.
- [65] "Special Issue on TerraSAR-X: Mission, Calibration, and First Results", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 2, February 2010.
- [66] Gray, A.L., Mattar, K.E., and Farris-Manning, P.J., "Airborne SAR Interferometry for Terrain Elevation," in *Proceedings IEEE International Geoscience and Remote Sensing Symposium*, vol. 2, pp. 1589–1591, 1992.
- [67] Livingstone, C.E., et al., "The Canadian Airborne R&D SAR Facility: The CCRS C/X SAR," in *Proceedings IEEE International Geoscience and Remote Sensing Symposium*, vol. 3, pp. 1621–1623, May 27–31, 1996.
- [68] Adams, G.F., et al., "The ERIM Interferometric SAR: InSARE," *IEEE AES Systems Magazine*, pp. 31–35, December 1996.
- [69] Mercer, J.B., Thornton, S., and Tennant, K., "Operational DEM Production from Airborne Interferometry and from RADARSAT Stereo Technologies," in *Proceedings of the American Society for Photogrammetry and Remote Sensing-Resource Technology, Inc. Conference*, Tampa, FL, March 31–April 3, 1998.
- [70] Wheeler, K. and Hensley, S., "The GeoSAR Airborne Mapping System," *Record of the IEEE International Radar Conference*, pp. 831–835, 2000.
- [71] Hensley, S., et al., "First P-Band Results Using the GeoSAR Mapping System," *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, vol. 1, pp. 126–128, 2001.
- [72] Amelung, F., et al., "Sensing the Ups and Downs of Las Vegas: InSAR Reveals Structural Control of Land Subsidence and Aquifer-System Deformation," *Geology*, vol. 27, no. 6, pp. 483–486, June 1999.
- [73] Wegmüller, U., et al., "Nonuniform Ground Motion Monitoring With TerraSAR-X Persistent Scatterer Interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 2, pp. 895–904, February 2010.

- [74] Romeiser, R. and Hirsch, O., “Possibilities and Limitations of Current Measurements by Airborne and Spaceborne Along-Track Interferometric SAR,” in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, vol. 1, pp. 575–577, 2001.
- [75] Wikipedia. “Northrop Grumman E-8 Joint STARS.” Available at http://en.wikipedia.org/wiki/Northrop_Grumman_E-8_Joint_STARS.
- [76] Gens, R. and Vangenderen, J.L., “SAR Interferometry—Issues, Techniques, Applications,” *International Journal of Remote Sensing*, vol. 17, no. 10, pp. 1803–1835, 1996.
- [77] Massonnet, D. and Feigl, K.L., “Radar Interferometry and Its Application to Changes in the Earth’s Surface,” *Rev. Geophys.*, vol. 36, no. 4, pp. 441–500, 1998.
- [78] Rogers, A.E.E. and Ingalls, R.P., “Venus: Mapping the Surface Reflectivity by Radar Interferometry,” *Science*, vol. 165, pp. 797–799, 1969.
- [79] Zisk, S.H., “A New Earth-Based Radar Technique for the Measurement of Lunar Topography,” *Moon*, vol. 4, pp. 296–300, 1972.
- [80] Graham, L.C., “Synthetic Interferometric Radar for Topographic Mapping,” *Proceedings of the IEEE*, vol. 62, pp. 763–768, June 1974.
- [81] Zebker, H.A. and Goldstein, R.M., “Topographic Mapping from Interferometric SAR Observations,” *Journal of Geophysics Res.*, vol. 91, pp. 4993–4999, 1986.
- [82] Li, F. and Goldstein, R.M., “Studies of Multibaseline Spaceborne Interferometric Synthetic Aperture Radars,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, pp. 88–97, 1990.
- [83] Richman, D., “Three Dimensional, Azimuth-Correcting Mapping Radar,” U.S. patent 4,321,601, March 23, 1982.
- [84] Eineder, M. et al., “Spaceborne Spotlight SAR Interferometry with TerraSAR-X,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1524–1535, May 2009.

8.12 | PROBLEMS

1. Compute the approximate number of posts (points at which an elevation measurement is specified) required to cover a region measuring $10 \text{ km} \times 10 \text{ km}$ for a DEM at DTED levels 0 through 5 and HRE levels 6 through 8.
2. Consider the sensor-scatterer geometry shown in Figure 8-4. Assume the center of the sensor baseline is at altitude $H = 2000 \text{ m}$, the ground range $y_1 = 5100 \text{ m}$, and the elevation of scatterer \mathbf{P}_1 is $h = 73.6 \text{ m}$. Also assume a sensor baseline length $B = 3 \text{ m}$ and a tilt angle $\beta = 90^\circ$ (vertically stacked sensors). Compute the depression angle ψ_1 from the baseline center to \mathbf{P}_1 . Next, compute the ranges R_a and R_b from each sensor to \mathbf{P}_1 . Then compute R and δR from R_a and R_b . Finally, use equations (8.5) and (8.2) to estimate ψ_1 and h and the percentage error in the estimate of each.
3. For the geometry of problem 2, compute the ground range y_0 at which the scatterer \mathbf{P}_1 will be imaged.
4. Continuing with the geometry of problem 2, the IPD measured at ground range y_0 for a 10 GHz radar ($\lambda = 3 \text{ cm}$) with independent transmitters ($p = 2$) is $\phi_{ab} = 444.0425$ radians. Use InSAR DEM algorithm 1 to estimate the height of the scatterer imaged at y_0 and its actual ground range; that is, estimate y_1 and $\hat{h}(y_1)$. Compute the percentage error in each.

5. Repeat problem 4 using InSAR DEM algorithm 2.
6. In [84] an example InSAR image is shown where the ambiguous height h_{amb} of the TerraSAR-X interferometric SAR is approximately equal to the height of the Eiffel Tower, which is 324 m. The collection conditions were $\lambda = 3.1$ cm, $p = 2$, $\psi = 34.7^\circ$, and an altitude of 514 km. Assuming an earth radius of 6371 km, the nominal range R to the earth's surface will be 637.79 km. Compute the perpendicular baseline B_\perp that must have been used to generate this image.
7. Compute the minimum SNR in dB needed to ensure that the coherence due to thermal noise is at least $\gamma_{noise} = 0.7$. Repeat for $\gamma_{noise} = 0.8$ and 0.9.
8. Compute the critical perpendicular baseline for each of the systems listed in Table 8-3. Use $p = 1$ for the one-pass systems and $p = 2$ for the two-pass systems and formation flying systems. Assume a nominal depression angle of 45° for the airborne systems and 60° for the spaceborne systems. Assume the range for the airborne systems is $\sqrt{2}$ times the altitude (due to the 45° depression angle); for the spaceborne systems, use the altitude as the range. Assume the ground range resolution is the same as the largest value given for the cross-range resolution (worst case for spatial coherence loss). For multifrequency systems, use the highest RF listed (worst case for coherence). Also determine the coherence factor $|\gamma_{spatial}|$ for each system under these conditions and the additional assumption that the perpendicular baseline B_\perp is the largest baseline value listed (worst case for spatial coherence loss). Repeat for the smallest resolution and the smallest B_\perp listed (best for spatial coherence), but do not change the RF.
9. Show that $|\hat{\gamma}|$ of equation (8.49) is equal to 1 when the two images are identical to within a scale factor, $a[n] = \alpha \cdot b[n]$ for some α .
10. Determine the maximum error in tilt angle β in degrees that can be tolerated if the error in height is limited to 1 m. Compute the tilt angle tolerance for a range of 20 km at a depression angle of 30° (airborne case) and again for 800 km and 45° (satellite case).

PART III

Array Processing and Interference Mitigation Technique

CHAPTER 9 Adaptive Digital Beamforming

CHAPTER 10 Clutter Suppression Using Space-Time Adaptive Processing

CHAPTER 11 Space-Time Coding for Active Antenna Systems

CHAPTER 12 Electronic Protection

Adaptive Digital Beamforming

David Aalfs

Chapter Outline

9.1	Introduction	401
9.2	Digital Beamforming Fundamentals	404
9.3	Adaptive Jammer Cancellation	419
9.4	Adaptive Beamformer Architectures	435
9.5	Wideband Cancellation	441
9.6	Summary	449
9.7	Further Reading	449
9.8	References	449
9.9	Problems	451

9.1 | INTRODUCTION

There are many excellent texts on adaptive array signal processing [1–8]. This chapter will not attempt to duplicate those efforts and will instead focus on the important aspects of applying digital beamforming (DBF) and adaptive techniques to radar. The development of phased array antennas in the 1950s was a critical step in the advent of multifunction radars. Electronic scanning for the first time provided sufficient beam agility to interleave search-and-track functions on multiple targets. A comparable advance in radar functionality is being realized in the next generation of phased array antennas by replacing analog combiner networks in the antenna with digital beamforming.

The benefits of DBF come at the expense of needing more receivers and higher computational throughput to perform operations digitally that were formerly done with analog hardware. Fortunately, as digital computing technology has continued to advance at a rapid rate, these more demanding computing requirements have become increasingly easier to accommodate.

Adaptive algorithms enable the radar system to be responsive to the signals received by the antenna by optimizing the weights in the digital beamformer with respect to some statistical optimization criterion, such as minimum mean square error (MMSE) or maximum signal-to-noise ratio (SNR). This allows unwanted interference to be automatically canceled while preserving gain in the direction of the target signals of interest.

9.1.1 Organization

The fundamentals of digital beamforming are covered in Section 9.2 including the key benefits and implementation challenges. Section 9.3 introduces adaptive algorithms for jammer cancellation and derives the optimal solution for the adaptive weights. Section 9.4 describes the sidelobe canceller, beamspace, and subarray space adaptive beamforming architectures and their relative performance. Section 9.5 extends the adaptive algorithms from section 9.3 to operate over wide bandwidths and presents the special case of linear frequency modulated (LFM) waveforms pulse compressed using stretch processing.

9.1.2 Key Points

- DBF enhances phased array radar functionality.
- Element level DBF is often impractical for large arrays operating at high frequencies.
- DBF at the subarray level reduces the receiver count for easier implementation.
- The subarray architecture choice heavily impacts performance due to grating lobes and grating nulls.
- ADBF increases the input-output (I/O) throughput requirements linearly with the number of spatial channels.
- ADBF increases the computational requirements superlinearly with the total number of both spatial and temporal degrees of freedom (DOF).
- Various types of spatial DOF can be used for adaptive cancellation, and they each work best for jamming located in specific regions of the antenna pattern.
- When canceling interference over wide bandwidths, the adaptive filter needs to jointly optimize over both the spatial and frequency domains.
- Cancellation performance is limited by the hardware errors in the array and receivers.

9.1.3 Notation

The following lists many of the variable names found within this chapter:

f_0 = center frequency (Hz)

ω_0 = center frequency (rads/s)

λ_0 = center wavelength (m)

M = number of channels

N = number of pulses

L = number of available range bins

T = pulse repetition interval (s)

c = velocity of propagation (speed of light, m/s)

\mathbf{s}_s = spatial steering vector

\mathbf{v}_s = hypothesized spatial steering vector

\mathbf{s}_t = temporal steering vector

\mathbf{v}_t = hypothesized temporal steering vector

\mathbf{s}_{s-t} = space-time steering vector

\mathbf{v}_{s-t} = hypothesized space-time steering vector

$\mathbf{x}_{s,k}(n)$ = spatial data snapshot, k -th range cell, n -th pulse

$\mathbf{x}_{t,k}(m)$ = temporal data snapshot, k -th range cell, m -th channel

\mathbf{x}_k = space-time data snapshot, k -th range cell

\mathbf{c}_k = clutter space-time snapshot, k -th range cell
 \mathbf{n}_k = uncorrelated noise space-time snapshot, k -th range cell
 \mathbf{t}_k = target space-time snapshot, k -th range cell
 \mathbf{R}_k = null-hypothesis covariance matrix, k -th range cell
 $\hat{\mathbf{R}}_k$ = null-hypothesis covariance estimate, k -th range cell
 \mathbf{w}_k = space-time weight vector
 $\hat{\mathbf{w}}_k$ = adaptive space-time weight vector
 $L_{s,1}$ = clairvoyant SINR loss
 $L_{s,2}$ = adaptive SINR loss
 P_d = probability of detection
 P_f = probability of false alarm
 σ_s^2 = single channel, single pulse signal power
 σ_n^2 = single channel, single pulse noise power
 σ_c^2 = single channel, single pulse clutter power
 P_s = output signal power
 P_n = output noise power

9.1.4 Acronyms

Acronyms used in this chapter include the following:

ADC	analog-to-digital converter
AGC	automatic gain control
COTS	commercial off-the-shelf
CPI	coherent processing interval
DBF	digital beamformer
DOF	degrees of freedom
EMI	electromagnetic interference
FFT	fast Fourier transform
FIR	finite impulse response
FPGA	field programmable gate array
GSC	generalized sidelobe canceller
I/Q	in-phase/quadrature
LCMV	linearly constrained minimum variance
LNA	low-noise amplifier
LO	local oscillator
MMSE	minimum mean square error
MVDR	minimum variance distortionless response
PRI	pulse repetition interval
PSD	power spectral density
RCS	radar cross section
RMB	Reed, Mallett, and Brennan
Rx	receiver
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
TVW	time-varying weights
Tx	transmitter
ULA	uniform linear array
WFG	waveform generator

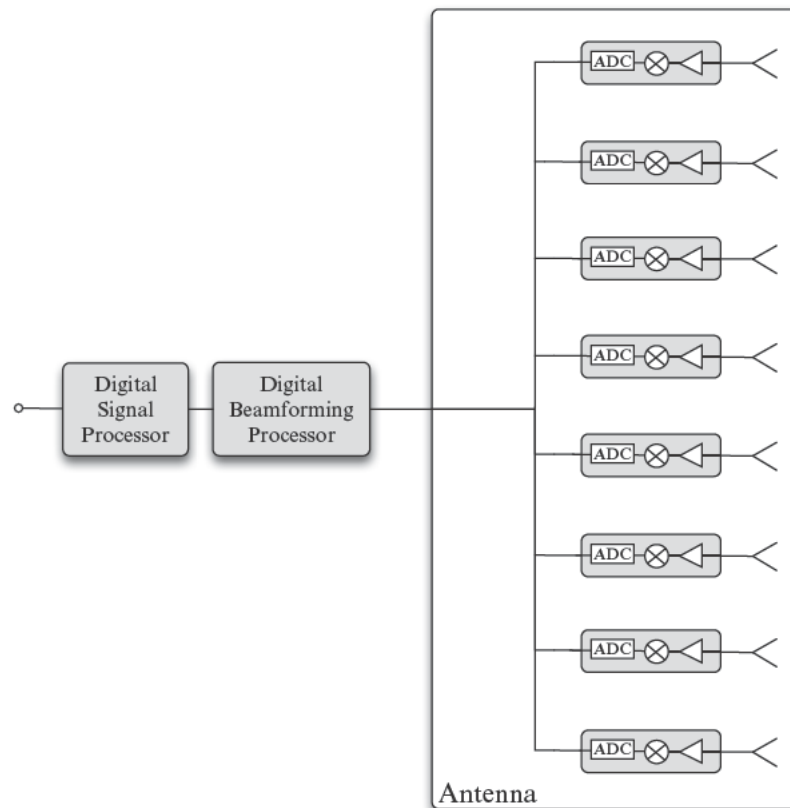
9.2 | DIGITAL BEAMFORMING FUNDAMENTALS

A conventional analog phased array antenna sums the signals from the individual antenna elements through radio frequency (RF) combiner networks made up of one or more stages of phase shifters, attenuators, amplifiers, and time delay units (for wideband systems). Only after the beams are fully formed in analog are receivers with analog-to-digital converters (ADCs) used to produce digital signals that are input to the signal and data processing computers for further processing. For more information on phased arrays, see [9]. In an antenna with a DBF architecture, the receivers are placed prior to the final stage of combining so that some or even all of the beamforming occurs in a digital computer.

Figure 9-1 shows a notional block diagram of an array that has element-level DBF on receive. When the array is in receive mode, the signals come in through the elements, pass through the low-noise amplifier (LNA), and then go directly into a receiver where the analog signal is converted into digital in-phase/quadrature (I/Q) samples. The digital I/Q samples are then sent to a digital beamforming processor where the beamforming is done in the computer.

Less commonly, DBF can also be done on transmit, where a reference signal is distributed to waveform generators (WFGs) at every element (or subarray). The WFGs produce the transmit waveform that goes through the high-power amplifier, the circulator, and then out through the element into free space. In a conventional non-DBF array there would also be a phase shifter and possibly a time-delay unit to steer the beam, but with

FIGURE 9-1 ■
Element-level digital
beamforming.



element-level DBF on transmit any necessary phase shifts or time delays can be performed in the distributed WFGs.

The DBF architecture enables the use of a number of array signal processing techniques that can enhance the capability of a radar, including the following:

1. Digital restearing of beams on receive for improved search occupancy
2. Adaptive cancellation for jammer and electromagnetic interference (EMI) mitigation
3. High-resolution angle estimation of both targets and jammers for improved metric accuracy

In addition, wideband DBF array designs can potentially replace costly analog time-delay units, which are typically a significant source of error, with effectively infinite precision digital time delay. A DBF array is also more robust to receiver failures because only the portion of the array feeding that receiver is lost as opposed to an entire sum or monopulse difference beam in a conventional analog array.

If there is a receiver at every element as shown in Figure 9-1, then any amplitude weighting used to control sidelobes and any phase weighting (or time delay for wideband) used to steer beams can be performed in the digital computer. Amplitude weights, phase weights, and time delays can be applied digitally with less error than can be achieved with analog components while greatly simplifying the front-end analog hardware.

The fundamental operation of the DBF processor is to weight and sum the I/Q samples from the spatial channels to produce beams. If the complex amplitude and phase weighting for the k -th element, x_k , is given by

$$w_k = a_k e^{j\phi_k} \quad (9.1)$$

then the beam output, y , is given by

$$y = \sum_{k=1}^M w_k^* x_k = \mathbf{w}^H \mathbf{x} \quad (9.2)$$

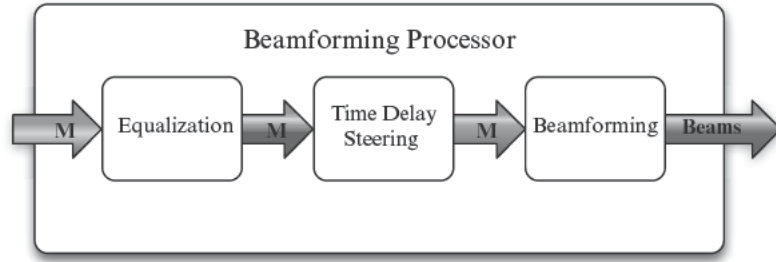
where

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix}, \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} \quad (9.3)$$

In addition to performing the weight and sum beamforming, the beamforming processor may also perform other functions such as equalization of the receiver channels and time-delay steering for wideband operation (see Figure 9-2). The receiver channel equalization is necessary to match their frequency responses over the bandwidth as closely as possible to enable good jammer cancellation.

Time-delay steering is necessary to prevent the beam from squinting in angle as a function of frequency during wideband operation. In a digital system, the simplest way to delay signals would be to shift the digitized data by multiples of the sampling period. Unfortunately for typical ADC sampling rates and radar operating frequencies, sample shifting does not provide enough granularity for time-delay steering a phased array. For example, ADC sampling rates from 10 to 100 MHz result in sampling period of 100 to 1 nsec. If the time delay needs to be done only at a subarray level, then the time delay can

FIGURE 9-2 ■
Typical functions of
a beamforming
processor



typically be quantized at a granularity of one to two wavelengths. Therefore, for radars operating at frequencies anywhere from 1 to 10 GHz, the required time delay granularity at the subarray level is 1 nsec to 0.1 nsec. Only at the higher ADC sampling rates and lower radar frequencies does sampling shifting alone suffice. Fortunately, subsample fractional delays can be implemented easily in digital hardware in many ways. One approach is to design finite impulse response (FIR) filters that give the desired fractional delays. The basic architecture of an FIR filter is shown in Figure 9-3, where Z^{-1} denotes a single sample delay and $(h_1 \dots h_n)$ are the filter coefficients [10]. Several design methods are available for deriving these filter coefficients (e.g., least squares, maximally flat, equiripple, and polyphase) [11]. Regardless of the design method, the filter coefficients $(h_1 \dots h_n)$ for all the required delays can be precomputed and then stored in the beamforming processor.

If an FIR filter in the beamforming processor is already being used for some other purpose (e.g., equalization), then it is possible to adjust the coefficients of that existing filter to create a new filter that also provides the desired amount of fractional delay, Δd [11]. The new filter coefficients are given by

$$\mathbf{h}_{new} = \mathbf{W}_{\Delta d} \mathbf{S}_{\Delta d} \mathbf{h}_{old} \quad (9.4)$$

where $\mathbf{W}_{\Delta d}$ is a diagonal matrix with a window function along the diagonal to control the sidelobes of the filter frequency response, and

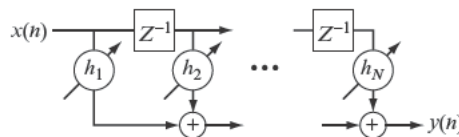
$$\mathbf{S}_{\Delta d, k, l} = \text{sinc}(k - l - \Delta d), \quad k, l = 1, 2, \dots, N. \quad (9.5)$$

9.2.1 DBF Implementation Challenges

The key challenges to implementing a DBF architecture are the following:

1. **Packaging:** Can the receivers be designed fit within the available physical area?
2. **Power:** Can the receiver and computing power consumption be kept within system constraints?
3. **Cost:** Can the receivers and the beamforming computer be built at an affordable price?
4. **Interference:** Can the receivers be made robust and sufficiently linear to handle strong interference?

FIGURE 9-3 ■
Fractional delay FIR
filter.



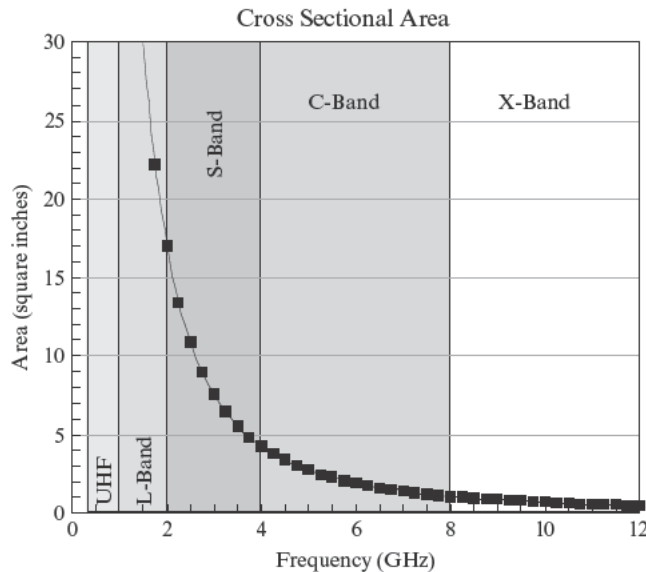


FIGURE 9-4 ■
Element cross sectional area decreases as frequency increases.

5. **Alignment and calibration:** Can the distributed receivers maintain adequate alignment and calibration?
6. **Reliability:** Can the receivers be made reliable enough to achieve adequate system availability?
7. **Data movement:** Can the digital data be communicated from the receivers to the beamforming computer in real time?
8. **Computation:** Can the beamforming processor keep up with its computational load?

The higher the frequency of the system, the smaller the cross sectional area behind the element available for packaging all the components needed to transmit and receive signals through that element (see Figure 9-4). If the DBF architecture requires that the receiver also fit within that cross sectional area, then at higher frequencies the DBF receivers must fit into ever smaller form factors. While it may be possible to use cables to fan out the RF signals to a larger cross sectional area per element, this adds cable losses, weight, and volume that may not be acceptable.

Radar receivers have historically been built in low quantities as custom hardware that is expensive and consumes a lot of power. When the number of receivers is small, as is the case for conventional RF beamforming, the cost and power consumption of the receivers, while significant on a per unit basis, is typically only a small portion of the overall system cost and power budget. Multiply that unit cost and power consumption by the number of receivers needed for DBF (e.g., tens, hundreds, or even thousands of units), and the impact at the system level can become significant. Therefore, for DBF to be feasible the receivers must be small, low cost, and low power.

Since one of the key benefits of DBF is to enable adaptive cancellation of strong interference, it is imperative that the DBF receivers be able to handle strong jamming signals without saturating or producing unwanted spurious signals due to nonlinear components in the receive chain. In conventional arrays where the receivers are located after all the beamforming has been done in analog hardware, the receiver is protected from strong side-lobe jamming signals by the low sidelobes of the full antenna pattern. In a DBF system, the

receivers are effectively seeing the world through a much smaller antenna that has less directivity. This means that receivers for DBF require higher dynamic range, better linearity, and better filtering because the interference has not been attenuated by the low sidelobes of a fully formed beam. To be able to cancel that jamming later on in signal processing, the receivers have to be able to adequately capture the usually weak signals of interest in the presence of strong unwanted interference without producing distortion in either. For more information on nonlinear distortion in DBF arrays and mitigation approaches see [12,13].

The distributed receivers also need to be kept in strict time alignment and matched to each other in terms of their frequency responses. The former is necessary to prevent gain loss and signal distortion in the beam output, and the latter is necessary to provide good jammer cancellation over the bandwidth. The mixers and the ADC converters in the receive chain require local oscillator (LO) signals and clocks, respectively, to perform their functions. Any misalignment of the LO and clock signals in the distributed receivers will result in a loss of signal coherence from channel to channel. Therefore, the network for distributing LO and clock signals to the receivers needs to be well calibrated and stable, which is an engineering challenge.

Receivers have a certain probability of failure that impacts the overall system availability. The increased number of receivers for DBF implies that the probability that any one of those receivers will fail will be higher. Availability is defined as the probability that a system or subsystem is operational at a given time. For simplicity, consider the impact of receiver failures on the system availability while neglecting all other contributors. If the availability of a single receiver channel is A_1 , then system availability for n receiver channels is given by

$$A_n = (A_1)^n \quad (9.6)$$

Since the single-channel availability is by definition less than 1, the system availability becomes much smaller as n becomes large. This assumes, however, that all the channels in a DBF system need to be available for the system to be considered available. That would be true for a monopulse radar ($n = 3$) where the loss of even a difference channel means that the radar can no longer provide angle estimates in one dimension. In a DBF system, each receiver channel represents a much smaller portion of the total aperture. For example, in a 16-channel DBF system the loss of 1/16 of the array due to a receiver failure results in less than 0.3 dB loss in gain. The higher the number of channels, the more gracefully performance degrades as channels fail. The availability for an N -channel DBF system that can tolerate one failed receiver is given by

$$A_{N:N-1} = A_1^N + NA_1^{N-1}(1 - A_1) \quad (9.7)$$

Figure 9-5 shows a plot of the system availability as a function of the single-channel availability for 3-channel monopulse, 16-channel DBF, 16-channel DBF with at least 15 channels available, and 16-channel DBF with at least 14 channels available. As expected, the 16-channel DBF availability is significantly reduced compared with the monopulse, but if the loss of one or two channels can be tolerated, then 16-channel DBF can provide higher availability than a monopulse system though at a degraded level of performance. Unfortunately, as the number of receiver channels increases, more and more receiver channel failures need to be tolerated for DBF to maintain availability equivalent to a 3-channel monopulse system. As a result, large-scale DBF is likely to require receivers with higher

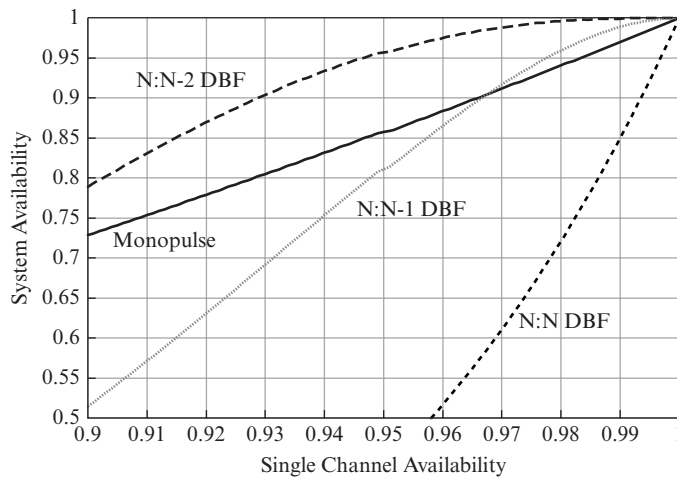


FIGURE 9-5 ■
System availability
for 16-channel DBF.

reliability or more rapid replacement with spares than conventional systems. Also, while the impact of the loss of a DBF receiver on the antenna gain may be relatively small, the impact on sidelobe levels can be more dramatic and should be taken into account for applications where maintaining low sidelobes is a high priority.

I/O and Computational Aspects

Once the array data have been converted into digital samples in the receivers, they need to be communicated to the beamforming computer. Figure 9-6 shows the I/O throughput required for 10-, 100-, and 1000-receiver channels as a function of the ADC rate assuming the number bits is 16 and that the I/Q samples are formed in the receiver. For large numbers of receivers and large sample rates, a beamforming processor implemented in commercial off-the-shelf (COTS) components, such as field programmable gate arrays (FPGAs) or graphical processing units (GPUs) is likely to be limited I/O capacity into the board and may require a large number of boards to support the necessary data rates.

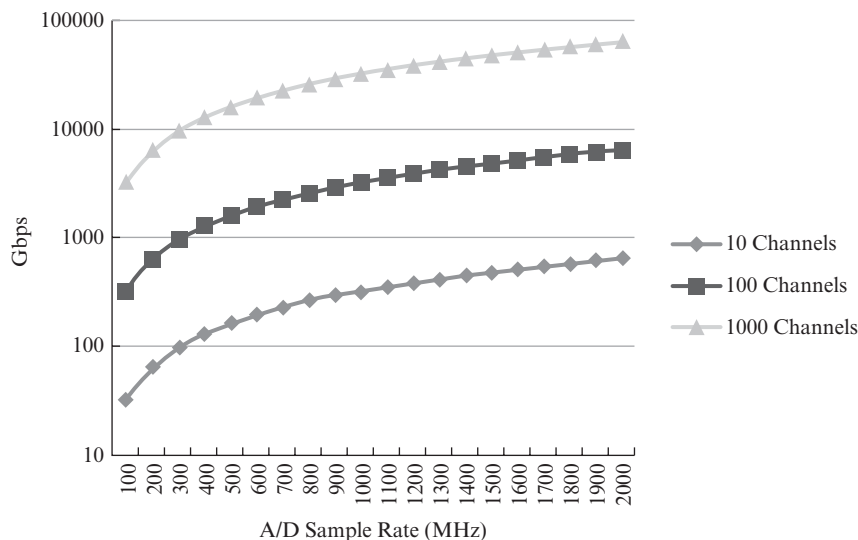


FIGURE 9-6 ■
I/O throughput in
Gbits/sec from the
receivers to the
beamforming
computer.

TABLE 9-1 ■ Number of Complex Multiplies and Adds for Adaptive Processing

Estimate covariance matrix	$O(N^2K)$
Invert covariance matrix	$O(N^3)$
Formulate adaptive weights	$O(MN^2)$
Apply adaptive weights	$O(MS)$

In the beamforming processor, the digital data can be combined into beams, subarrays, auxiliary channels, or combinations of these as dictated by the choice of adaptive beamforming architecture (e.g., beamspace, subarray space, sidelobe canceller). Any non-adaptive combining that precedes the adaptive beamforming involves relatively simple vector–matrix multiplies that are highly parallelizable and can be efficiently implemented on FPGA or GPU hardware.

Adaptive beamforming is discussed in more detail in Section 9.3, but the computational load can generally be broken down into four basic steps:

1. Estimate a covariance matrix.
2. Invert the covariance matrix.
3. Formulate the weights.
4. Apply the weights to the data.

The covariance matrix is estimated from the spatial channels being used for adaption and is numerically inverted. Next, some matrix–vector multiplies formulate the adaptive weights, and finally the weights are applied to the data to form beams. The number of complex operations (1 complex multiply/add) for each of these steps are shown in Table 9-1, where

N = number of spatial channels

S = number of samples in waveform

K = number of samples for covariance matrix ($K \leq S$, $K \geq 10N$)

M = number of simultaneous beams (three for monopulse)

For shorter pulses and larger numbers of DBF channels the computational load will be dominated by the covariance matrix inverse, and for longer pulses with fewer DBF channels the computational load will be dominated by the application of the weights to the data.

All of these implementation challenges are made easier if the receiver channel count is reduced. The trend for implementing DBF into radar systems has been to start at the lower frequency bands and progress into the higher bands over time. At the lower frequencies, the larger cross sectional areas and lower bandwidths make it easier to build low-cost receivers, and the lower data rates reduce the I/O throughput required from the receivers to the beamforming computer. The other trend has been to start with one-dimensional digital beamforming, usually in elevation, and progress toward two-dimensional digital beamforming, usually at a subarray level. Restricting the DBF to a single dimension and employing conventional RF analog beamforming in the other dimension keeps the receiver count to a manageable level while still providing the advantages of DBF, albeit in that one dimension only. This approach has been adopted for surveillance radars that rotate in azimuth and do digital beamforming in elevation. The SMART-L surveillance radar built by THALES (shown in Figure 9-7) is an example of this type of system. The MESAR 2



FIGURE 9-7 ■
SMART-L surveillance radar built by THALES has one-dimensional element-level digital beamforming in elevation (Picture taken at the Kiel Week 2007 onboard F221 Hessen of the German Navy. Released into the public domain by the author. <http://en.wikipedia.org/wiki/File:SmartLR.jpg>).

built by BAE Systems INSYTE is an example of a two-dimensional digital beamforming system at S band.

9.2.2 Subarrays

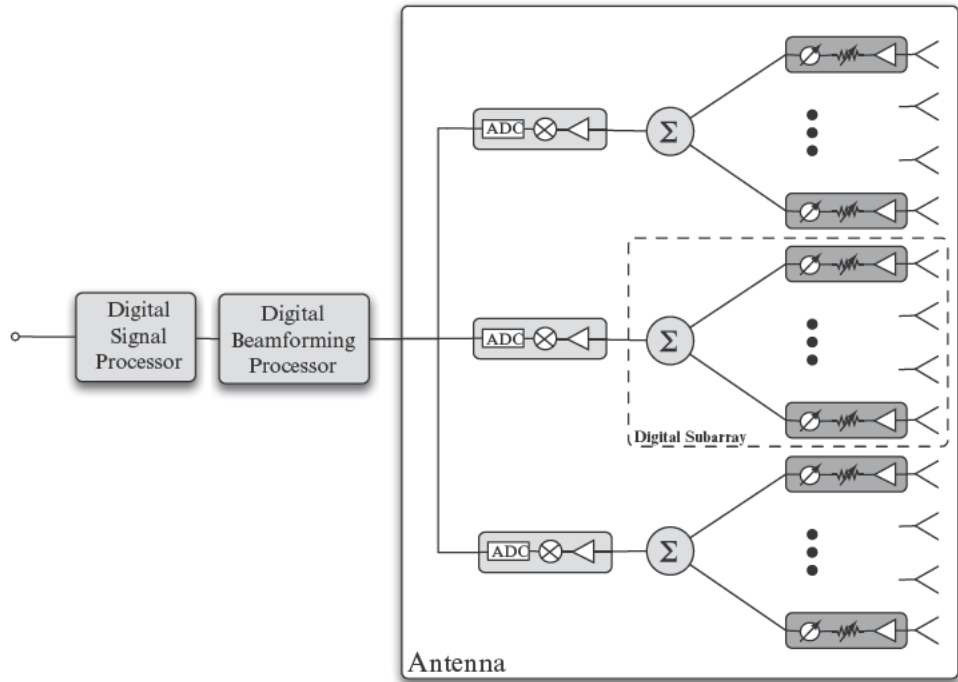
Given all the implementation challenges discussed in Section 9.2.1, DBF at the element level is often not practical for large phased arrays operating at high frequencies. As an alternative, a hybrid scheme of analog beamforming into subarrays prior to feeding into the digital receivers can be used to reduce the receiver count as shown in Figure 9-9. Analog attenuators and phase shifters in the transmit/receive (T/R) module provide the element-level control of the aperture taper weights and phase steering necessary to maintain good sidelobe and beamsteering performance. If the array is designed to support wideband waveforms, then time delay can be done digitally at the subarray level to mitigate beam squint over the full waveform bandwidth. Within the digital computer, complex weights can be applied to the subarray data to form multiple receive beams, to perform adaptive interference cancellation, or to implement some other form of spatial filtering in the same way it is done for element-level DBF.

A subarray mapping that assigns the M elements to N digital subarrays can be described by an $N \times M$ partition matrix \mathbf{P}_{dsa} . If the digital subarrays are further combined



FIGURE 9-8 ■
MESAR 2 radar built by BAE Systems INSYTE has two-dimensional subarray-level digital beamforming. (Courtesy of BAE Systems. With Permission.)

FIGURE 9-9 ■
Subarray-level
digital beamforming.



into a smaller number of adaptive subarrays, then that mapping can also be described by another partition matrix \mathbf{P}_{ssa} . Given these two stages of subarray combining, the beamforming equation can be rewritten as

$$\mathbf{y} = \mathbf{w}_a^H \mathbf{P}_{ssa} (\mathbf{w}_{beam}^H \mathbf{P}_{dsa} \mathbf{x}) \quad (9.8)$$

where \mathbf{w}_a is an adaptive weighting that is applied at the super subarray level, and \mathbf{w}_{beam} is a phase-only weighting at the digital subarray level used to form offset beams.

The choice of the subarray layout has a strong impact on the performance of the array signal processing algorithms that are applied at the subarray level. The design of subarrays has been analyzed previously in [14–20]. Those papers focus on optimizing the subarrays with respect to adaptive interference cancellation performance. In a multifunction radar, other performance metrics, such as offset beam sidelobe level, are also of interest. Using these other performance metrics as the criterion for optimizing the subarray layout can lead to different optimal solutions [21].

The most common candidates for the subarray architecture are regular, irregular, or overlapped designs. These subarray architecture options are illustrated in Figure 9-10 for the one-dimensional case, and they can be easily extended to two-dimensional arrays. A regular subarray architecture is the simplest to design and manufacture because all the subarrays are identical, but it also has the worst antenna pattern artifacts due to periodic grating lobes. One way to break up the grating lobes is to make the subarrays irregular such that their phase centers are not periodically distributed across the array, and each of the subarrays has a different antenna pattern response. The irregular subarray architecture breaks up the grating lobes, but it does so by pushing down grating lobe peaks while increasing the sidelobes everywhere else (filling in the valleys). As a result, irregular

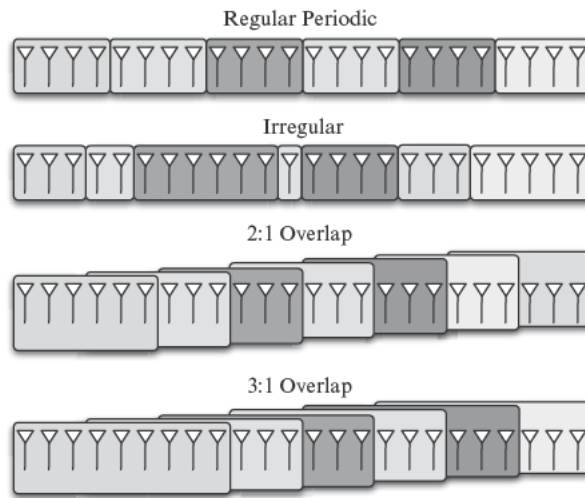


FIGURE 9-10 ■
Subarray
architecture options.

subarrays typically improve the peak sidelobe levels but don't improve or even degrade the average or root mean square (RMS) sidelobe levels.

An overlapped subarray architecture is attractive for applications that require multiple beams on receive because it has very good sidelobe performance in the offset beams [22]. As the name suggests, the subarrays overlap such that each element, except for elements at the edges of the array, contributes to multiple subarrays. If the subarrays have a 3:1 (or 75%) overlap, then elements are shared by three subarrays for a one-dimensional array or nine subarrays for a two-dimensional array. This architecture has the disadvantage of increased cost, complexity, and losses associated with manufacturing of the overlapped combining networks.

To understand how these different types of subarray layouts affect the resulting antenna pattern sidelobe performance, it helps to think of the antenna pattern as the product of three terms: the element pattern, the subarray pattern, and the array factor as shown in Figure 9-11. The case depicted in Figure 9-11 is for a regular subarray architecture with uniform weights at both the element and subarray levels. Weightings applied at the subarray level (e.g., for offset beamsteering or adaptive cancellation) shape the array factor producing grating lobes and nulls that exhibit the same periodicity as the array factor. An irregular subarray architecture will generally have subarray patterns that differ from each other, and more importantly it breaks up the periodicity of the array factor. This has

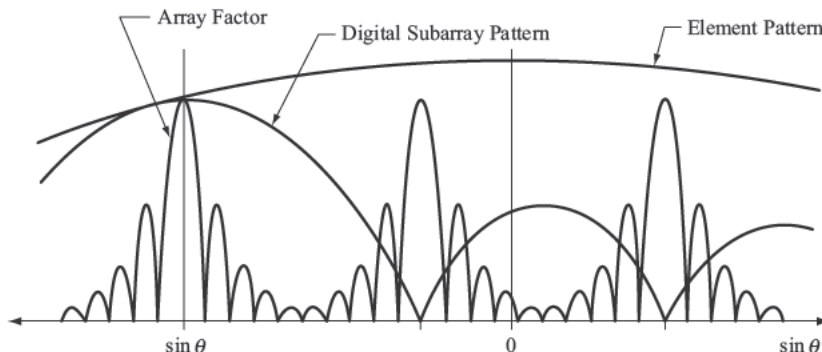
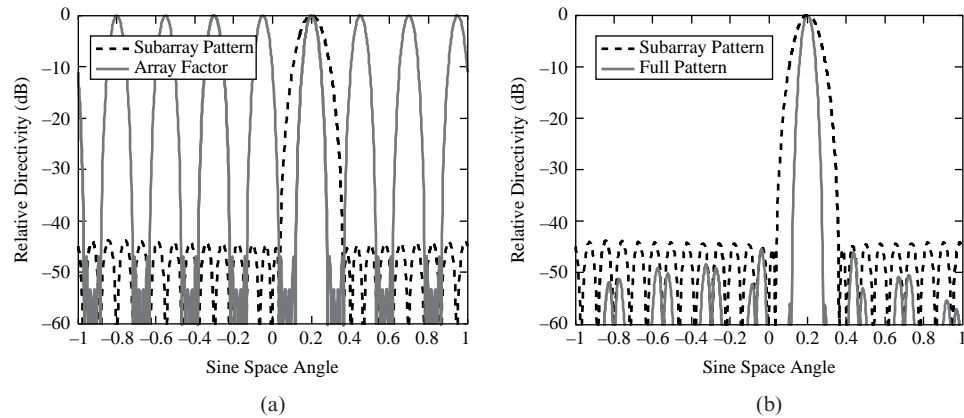


FIGURE 9-11 ■
Relationship
between the array
factor and the
subarray pattern.

FIGURE 9-12 ■
Array factor and subarray pattern for a 3:1 overlapped subarray architecture.



the effect of reducing the peak sidelobes and filling in the grating nulls. The overlapped architecture applies low sidelobe weighting both within the subarray and at the subarray outputs, which causes the both subarray pattern and the array factor to have low sidelobes, as shown in Figure 9-12. This results in a composite antenna pattern that has very low sidelobes and is more robust to perturbations in the subarray-level weighting.

9.2.3 Benefits of Digital Beamforming

The primary motivation for implementing DBF in a phased array radar is to provide multiple spatial channels in the digital computer for advanced signal processing algorithms to exploit. Enhanced functionality enabled by the DBF architecture can include the following:

- Adaptive nulling of jammers
- Multiple simultaneous beams on receive
- High-resolution angle estimation
- Enhanced dynamic range
- Reduced phase noise on transmit
- Digital time delay on transmit or receive
- Array healing

Most of this chapter focuses on adaptive jammer cancellation, but this section will briefly describe forming multiple simultaneous beams with a DBF architecture.

9.2.4 Multiple Simultaneous Beams on Receive

Once the subarray data are available in the digital computer, they can then be processed in multiple ways without impacting radar resources. For example, by applying the appropriate phase weightings to the data from each subarray, beams that are offset in angle from the original beam location used to set the element-level phase shifters can be synthesized in the digital computer.

For a uniform linear array (ULA) with the subarray architecture shown in Figure 9-13, the beam pattern response as a function of the sine space angle coordinate and wavelength

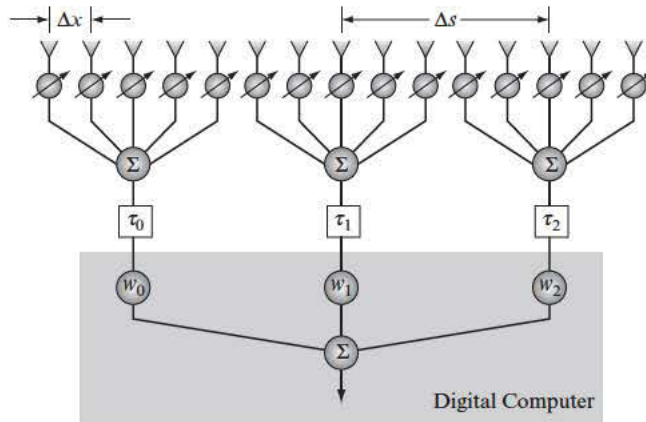


FIGURE 9-13 ■ Subarray architecture with element-level phase steering, subarray-level time-delay steering, and subarray-level offset beamsteering.

is given by

$$F(u, \lambda) = E_e(u) \sum_{m=0}^{M-1} W_m e^{-j(\frac{2\pi}{\lambda})m\Delta s(u-u_0)} \sum_{n=0}^{N-1} w_{mn} e^{-j2\pi n\Delta x(\frac{u}{\lambda} - \frac{u_0}{\lambda_0})} \quad (9.9)$$

where $u_0 = \sin(\theta_0)$ is the steered direction of the mainbeam in sine space, $E_e(u)$ is the element pattern, N is the number of elements per subarray, w_{mn} is the element-level amplitude weighting, Δx is the element spacing, M is the number of subarrays, Δs is the spacing between subarrays, and W_m allows for additional amplitude weighting at the subarray level.

For a uniform linear array of omnidirectional elements and uniform amplitude weighting, the beampattern response reduces to

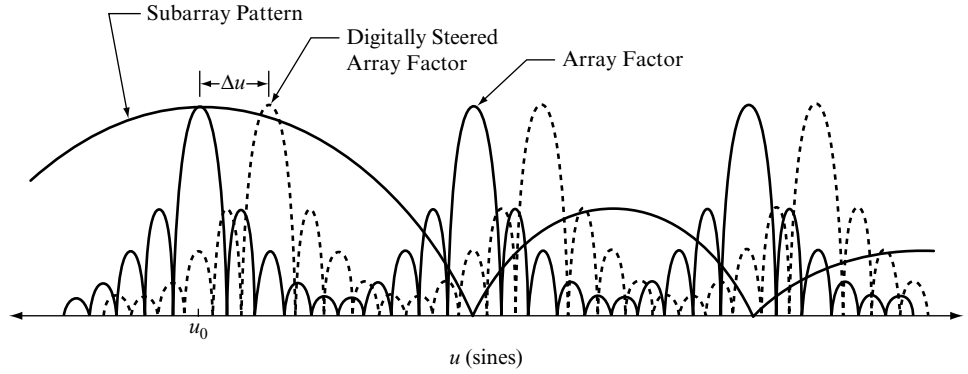
$$|F(u, \lambda)| = \left| \frac{\sin \left[\frac{\pi}{\lambda} M \Delta s (u - u_0) \right]}{\sin \left[\frac{\pi}{\lambda} \Delta s (u - u_0) \right]} \cdot \frac{\sin \left[N \pi \Delta x \left(\frac{u}{\lambda} - \frac{u_0}{\lambda_0} \right) \right]}{\sin \left[\pi \Delta x \left(\frac{u}{\lambda} - \frac{u_0}{\lambda_0} \right) \right]} \right| \quad (9.10)$$

Note that the pattern response is given by the product two periodic sinc functions. The first sinc term corresponds to the array factor, and the second sinc term corresponds to the subarray pattern. Figure 9-11 shows the array factor and subarray pattern for a typical case where the array has been steered to u_0 . For wideband operation, the subarray pattern squints in angle as a function of frequency, but the array factor will remain fixed if time delay steering is provided at the subarray level. Digitally restearing the beam at the subarray level manifests as a scanning of the array factor, as illustrated in Figure 9-14. In the figure, the array factor has been resteeered to $u_0 + \Delta u$, but the fact that the subarray pattern remains steered to u_0 causes the following problems:

1. Roll-off of the subarray pattern causes a gain loss in the resteeered direction.
2. The nulls of the subarray pattern no longer fall on peaks of the array factor, producing grating lobes in the full beampattern.
3. The peak of the full beampattern may not be located exactly at $u_0 + \Delta u$.

Consider an array that is steered to a sine space angle u_0 with a beampattern given by (9.10). To digitally resteer the array to a new location $u_0 + \Delta u$, a natural approach might be to digitally produce an incremental amount of time delay such that the resulting

FIGURE 9-14 ■
Array factor and subarray pattern for a resteeered beam.

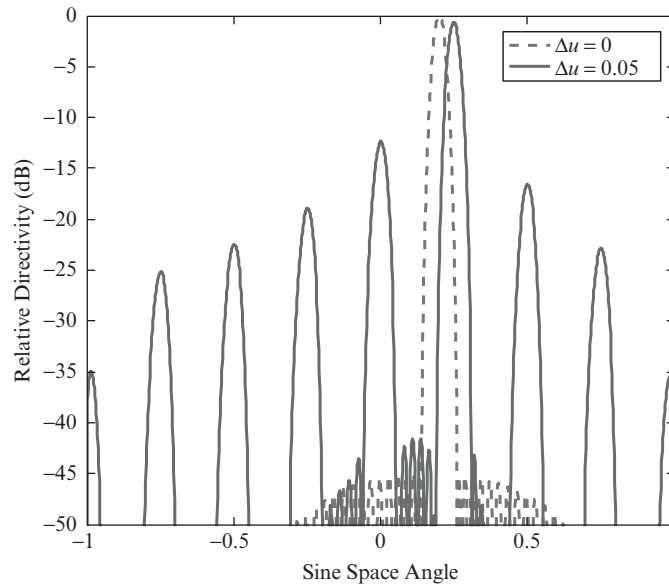


beam pattern is given by

$$|F(u, \lambda)| = \left| \frac{\sin \left[\frac{\pi}{\lambda} M \Delta_s (u - u_0 - \Delta u) \right]}{\sin \left[\frac{\pi}{\lambda} \Delta_s (u - u_0 - \Delta u) \right]} \cdot \frac{\sin \left[N \pi \Delta_X \left(\frac{u}{\lambda} - \frac{u_0}{\lambda_0} \right) \right]}{\sin \left[\pi \Delta_X \left(\frac{u}{\lambda} - \frac{u_0}{\lambda_0} \right) \right]} \right| \quad (9.11)$$

Figure 9-15 shows an example of a digitally resteeered antenna pattern for a regular subarray layout where $M = 8$, $N = 8$, $u_0 = 0.2$ sines, and $\Delta u = 0.05$ sines, and a Taylor weighting has been applied to provide low nominal sidelobes. The dashed curve is the pattern for the beam fully steered to 0.2 sines, and the solid curve is the pattern resteeered at the subarray level from 0.2 sines to 0.25 sines by scanning the array factor to $u_0 + \Delta u$. The directivity of the resteeered beam has dropped by about 0.65 dB relative to the fully steered beam due to the roll-off of the subarray pattern, and grating lobes occur at an interval of approximately 0.25 sines due to the periodicity of the array factor. The loss in directivity is inherent in subarray level resteeering and cannot be avoided, but the grating lobes can be mitigated by using either irregular or overlapped subarrays.

FIGURE 9-15 ■
Example of a digitally resteeered beam for a regular periodic subarray architecture.



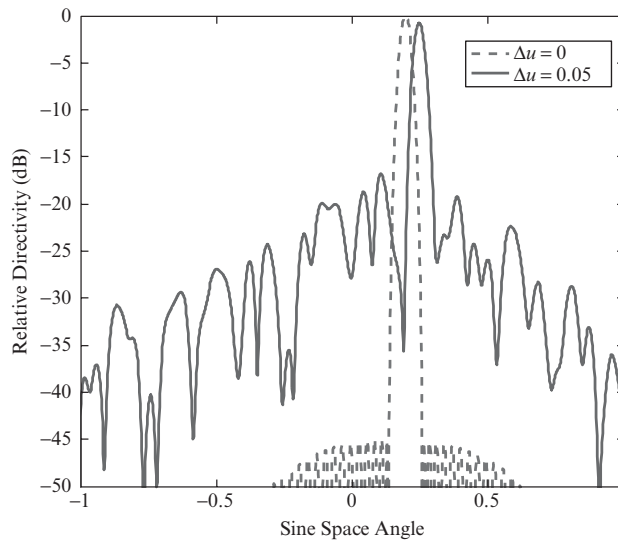


FIGURE 9-16 ■
Example of a
digitally resteeered
beam for an irregular
subarray
architecture.

If the subarrays are arranged in an irregular layout, then the grating lobes can be suppressed by breaking up the periodicity of the array factor. Figure 9-16 shows a digitally resteeered pattern for the same example 64-element array combined into 8 irregular subarrays. The peak sidelobe levels in the resteeered beam have been reduced, but the average sidelobe levels remain relatively high. Also, each of the subarrays in an irregular architecture will typically have different subarray pattern shapes and different directivity when the number of elements per subarray is allowed to vary. In environments with strong interfering signals such as clutter and jamming, an irregular subarray architecture that has highly dissimilar subarray patterns may have to apply automatic gain control (AGC) or suffer saturation in some subarrays but not in others complicating both the receiver design and the subsequent signal processing.

An overlapped subarray architecture provides the best sidelobe performance for digitally resteeered beams. Figure 9-17 shows a digitally resteeered pattern for a 64-element

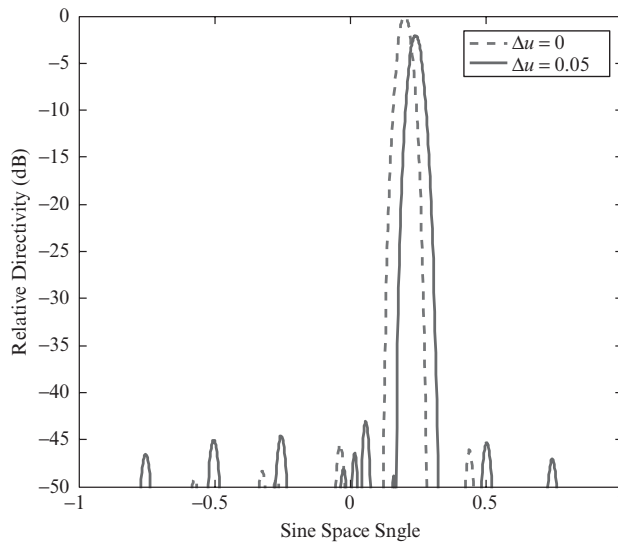
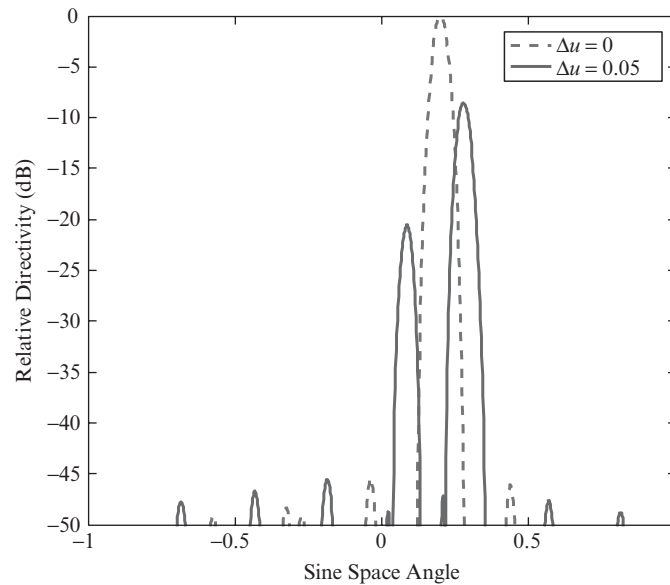


FIGURE 9-17 ■
Example of a
digitally resteeered
beam for a 3:1
overlapped subarray
architecture.

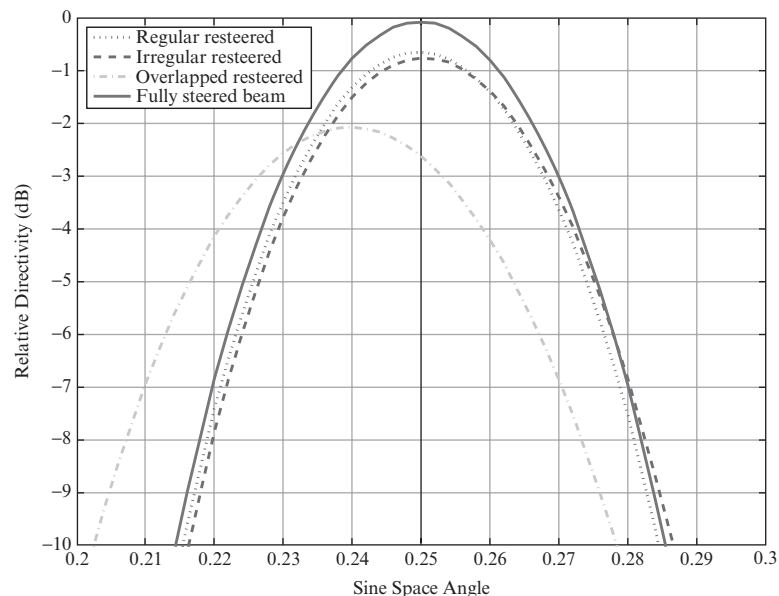
FIGURE 9-18 ■
Example of digitally resteeering beyond the point where the overlapped subarray architecture can maintain low sidelobes.



array combined into six 3:1 overlapped subarrays, where each subarray is composed of 24 elements. Low sidelobes in resteeered beams are maintained as long as the beam isn't resteeered to the point where the next lobe of the array factor begins to enter the mainlobe of the subarray pattern, as shown in Figure 9-18. The increase in the subarray size due to overlapping results in subarray patterns that roll off more sharply, leading to greater directivity loss as a function of the amount of resteeer. This directivity loss is noticeable in Figure 9-17 and blatant in Figure 9-18.

The degree to which the peak of the resteeered beam is offset from the desired location, $u_0 + \Delta u$, is determined by the shape of the subarray pattern and the amount of resteeer. Figure 9-19 shows resteeered beams for the regular, irregular, and overlapped subarray

FIGURE 9-19 ■
Subarray-level digitally resteeered beams compared with a fully steered beam.



architectures compared with a beam fully steered to $u_0 + \Delta u$. The regular and overlapped subarray architectures have a relatively small amount of directivity loss and almost negligible beam pointing error because the average subarray pattern is fairly broad compared with the amount of recenter. However, the overlapped architecture has a significant beam-pointing error in addition to the greater directivity loss. Both of these effects are due to the narrower subarray pattern, but the beam-pointing error can be eliminated by applying a precomputed correction factor for each desired offset angle, Δu .

9.3 | ADAPTIVE JAMMER CANCELLATION

Sidelobe jamming can be suppressed by designing antennas to have very low sidelobes. Unfortunately, the degree to which the sidelobes can be reduced using open-loop pattern synthesis techniques is fundamentally limited by the amplitude and phase errors in the array. Additional suppression beyond the error floor of the antenna can be achieved by employing adaptive cancellation methods. They allow the antenna to be responsive to the signal environment in which it operates by automatically adjusting the beamformer weights to minimize the jammer power at the output of the beamformer [23].

To understand how an adaptive digital beamformer functions, consider the linear array model depicted in Figure 9-20. In the figure, M antenna elements are uniformly distributed along a line with a separation of d units. The signals at each element pass through a receive chain that adds thermal noise, which is assumed to be independent from channel to channel. The outputs of the M receivers form an array data vector, \mathbf{x} , that is multiplied by the weight vector, \mathbf{w} , to form the beamformed output y .

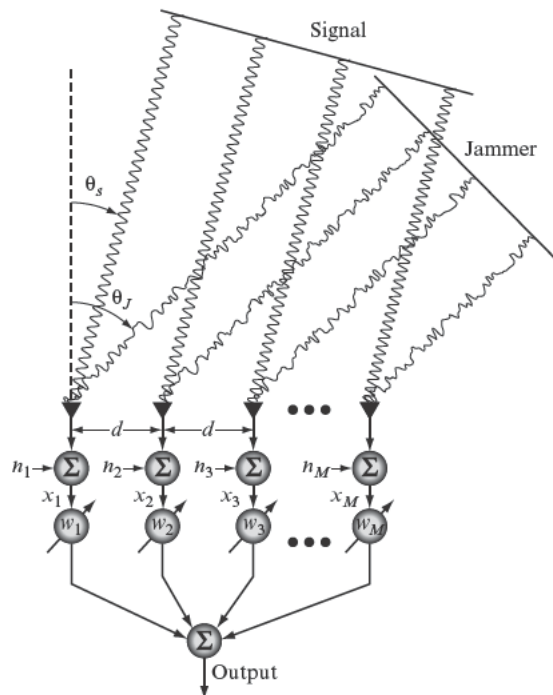
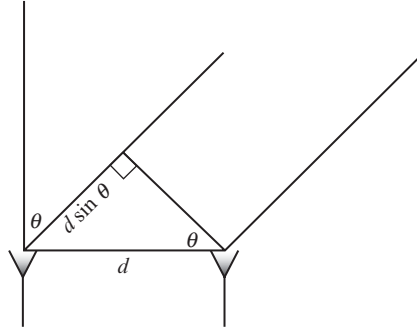


FIGURE 9-20 ■ Uniform linear array model.

FIGURE 9-21 ■
Propagation delay.



In vector notation, the beamformer output can be written as the inner product of the weight vector and the array data vector

$$y = \mathbf{w}^H \mathbf{x} \quad (9.12)$$

where the array data matrix, \mathbf{x} , is composed of signals due to sources that are external to the array such as targets and jammers, and thermal noise due to active components in the receive chain.

Assuming that the external sources are in the far field of the antenna pattern, they produce plane waves that impinge upon the array at an angle corresponding to the direction of arrival of the source. Thus, the signals in the digital beamformer channels due to plane waves are delayed by a factor of $d \sin \theta / c$ from channel to channel as shown in Figure 9-21.

The phase relationship from channel to channel is described by the steering vector and is a function of the element spacings, the angle of arrival, and frequency. The steering vector for the linear array in Figure 9-20 is given by

$$\mathbf{v}(\theta) = \left[1 \ e^{j2\pi \left(\frac{d}{\lambda}\right) \sin(\theta)} \ \dots \ e^{j2\pi (M-1) \left(\frac{d}{\lambda}\right) \sin(\theta)} \right]^T \quad (9.13)$$

where all the phases have been measured with respect to the first element. Once the steering vector is known, the array data vector at time t can be written as the sum of signal, jammer, and noise terms:

$$\mathbf{w}^H \mathbf{v}(\theta_i) = 0, \quad i = 1, 2, \dots, K \quad (9.14)$$

where $s(t)$ is the complex envelope of the target signal, $a_k(t)$ is the complex envelope of the k -th jammer, and \mathbf{n} is the noise vector.

The fundamental problem in adaptive beamforming is to find a weight vector, \mathbf{w} , based only on observing the total array data vector, \mathbf{x} , that minimizes the jammer term without degrading the signal term. If the jammers are located at the angles $(\theta_1, \theta_2, \dots, \theta_K)$, then to perfectly null all the jammers, \mathbf{w} might ideally be chosen such that

$$\mathbf{w}^H \mathbf{v}(\theta_i) = 0 \quad i = 1, 2, \dots, K \quad (9.15)$$

This has the geometric interpretation that the weight vector is perpendicular to all of the jammer steering vectors as shown in Figure 9-22. While a weight vector selected in that manner, if it exists, would have a perfect null in the direction of every jammer, there

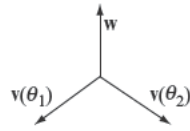


FIGURE 9-22 ■ Perfect nulls are formed when the weight vector is perpendicular to all the jammer steering vectors.

is no guarantee that it will also preserve the target signal. This approach also relies on knowledge of the jammer locations, and errors in their estimated locations would put the nulls in the wrong place.

Adaptive algorithms determine the best solution for the weights by performing an optimization with respect to selected statistical performance criteria. Figure 9-23 shows a block diagram of a generic optimum filter. The optimum filter takes a vector of inputs and combines them in some way to produce an output $y(n)$. The difference between the actual filter output and the desired output signal, $d(n)$, results in an error signal that the optimum filter is designed to minimize. The choice of statistical optimization criterion determines the form of the optimum filter.

Most adaptive beamforming algorithms are based on optimizing the weight vector with respect to the power of various signals after beamforming, such as minimizing the mean square error (MSE) signal or maximizing the signal power relative to the jammer-plus-noise power. The output power of the beamformer is given by

$$P_{out} = E[|y|^2] = \mathbf{w}^H E[\mathbf{x}\mathbf{x}^H] \mathbf{w} = \mathbf{w}^H R_x \mathbf{w} \quad (9.16)$$

where $E[\cdot]$ denotes the expected value, and R_x is the covariance matrix. Assuming the signal, jamming, and noise are mutually uncorrelated, the covariance matrix can be written as

$$\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^H] = \underbrace{|S|^2 \mathbf{v}(\theta_s) \mathbf{v}^H(\theta_s)}_{R_s} + \underbrace{\sum_{k=1}^K |a_k|^2 \mathbf{v}(\theta_k) \mathbf{v}^H(\theta_k)}_{R_j} + \underbrace{\sigma_n^2 \mathbf{I}}_{R_n} \quad (9.17)$$

As a simple example, consider a data matrix consisting of a single plane wave plus noise,

$$\mathbf{x} = S(t) \mathbf{v}(\theta) + \mathbf{n} \quad (9.18)$$

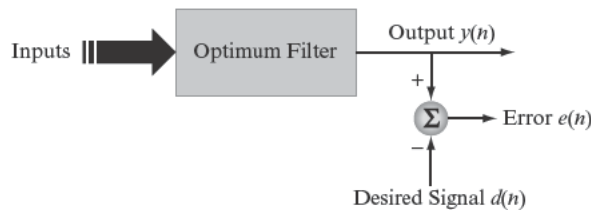


FIGURE 9-23 ■ Diagram of a generic optimum filter.

Then the covariance matrix has the form

$$\mathbf{R}_x = \begin{bmatrix} |S|^2 + \sigma_n^2 & |S|^2 e^{j2\pi f_0 \frac{d \sin \theta}{c}} & \dots & |S|^2 e^{j2\pi f_0 (M-1) \frac{d \sin \theta}{c}} \\ |S|^2 e^{-j2\pi f_0 \frac{d \sin \theta}{c}} & |S|^2 + \sigma_n^2 & \dots & |S|^2 e^{j2\pi f_0 (M-2) \frac{d \sin \theta}{c}} \\ \vdots & \vdots & \ddots & \vdots \\ |S|^2 e^{-j2\pi f_0 (M-1) \frac{d \sin \theta}{c}} & |S|^2 e^{-j2\pi f_0 (M-2) \frac{d \sin \theta}{c}} & \dots & |S|^2 + \sigma_n^2 \end{bmatrix} \quad (9.19)$$

where σ_n^2 is the noise power. Notice that the noise contributes only along the diagonal of the covariance matrix. This is because the diagonal elements of the covariance matrix correspond to the autocovariance of each spatial channel, which is equal to the sum of the signal power and the noise power. The off-diagonal elements of the covariance matrix are due to plane wave sources and contain information about the angle of arrival. It is the information in these off-diagonal terms that is exploited by adaptive algorithms to cancel jammers and by high-resolution angle estimation algorithms to locate sources.

9.3.1 Wiener Filter

If the optimization criterion for the adaptive filter is to minimize the mean square error, then the resulting solution for the adaptive weights is known as the Wiener filter. Recall that the error at the output of the beamformer can be written as

$$\varepsilon(t) = d(t) - y(t) = d(t) - \mathbf{w}^H \mathbf{x}(t) \quad (9.20)$$

where $d(t)$ is a reference signal that is highly correlated with the target signal of interest. The MSE is given by

$$\text{MSE} = E[|\varepsilon(t)|^2] = |d(t)|^2 - \mathbf{w}^H \mathbf{r}_{dx} - \mathbf{r}_{dx}^H \mathbf{w} + \mathbf{w}^H \mathbf{R}_x \mathbf{w} \quad (9.21)$$

where \mathbf{R}_x is the covariance matrix and

$$\mathbf{r}_{dx} = E[d^*(t) \mathbf{x}(t)] \quad (9.22)$$

is the cross-correlation vector between the desired signal and the array data vector. Minimizing the MSE with respect to the weight vector results in an optimal weight vector given by

$$\mathbf{w}_{opt} = \mathbf{R}_x^{-1} \mathbf{r}_{dx} \quad (9.23)$$

This form of the Wiener filter is not typically used for radar applications because it will attempt to form beams in the direction of any signals that are correlated with the desired signal. This can result in highly distorted beampatterns when there is specular multipath or coherent jamming present. To eliminate the need for the desired signal, consider the ideal case where the desired signal is set exactly equal to the target signal, $s(t)$. Then assuming that the target signal is uncorrelated with the jamming and the noise, the cross-correlation vector reduces to

$$\mathbf{r}_{dx} = E[S^*(t) \mathbf{x}(t)] = |S|^2 \mathbf{v}(\theta_s) \quad (9.24)$$

and plugging this into the optimum weight equation gives

$$\mathbf{w}_{opt} = |S|^2 \mathbf{R}_x^{-1} \mathbf{v}(\theta_s) \quad (9.25)$$

So far, this version of the Wiener filter does not appear to have improved matters because the need to know the desired signal has been replaced only with a need to know the target signal power and angle of arrival. However, while a radar may not know a priori where the target signal is located, it does know where it is supposed to point the beam. There may or may not be a target at that location, but if θ_s is replaced by the commanded beam location, θ_0 , then the adaptive weights will try to maintain gain in the commanded beam location while cancelling signals originating from any other direction. This results in the following form of the Wiener filter:

$$\mathbf{w}_{opt} = k \mathbf{R}_x^{-1} \mathbf{v}(\theta_0) \quad (9.26)$$

where the signal power, $|s|^2$, has been replaced with an arbitrary, nonzero scaling constant k because it has no impact on the output signal-to-interference-plus-noise ratio (SINR).

9.3.2 Maximum SINR

A more familiar choice of optimization criterion for the radar engineer is to maximize the SNR or SINR. The signal power, P_s , and the interference plus noise power, P_{i+n} , are given by

$$P_s = |\mathbf{w}^H \mathbf{s} \mathbf{v}(\theta)|^2 = |s|^2 \mathbf{w}^H \mathbf{v}(\theta) \mathbf{v}^H(\theta) \mathbf{w} = \mathbf{w}^H \mathbf{R}_s \mathbf{w} \quad (9.27)$$

$$P_{i+n} = \mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w} \quad (9.28)$$

and therefore by taking the ratio the SINR is equal to

$$\frac{P_s}{P_{i+n}} = \frac{|\mathbf{w}^H \mathbf{s} \mathbf{v}(\theta)|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} \quad (9.29)$$

The optimum weight to maximize SINR can be found using a trick involving the Schwarz inequality. The first step is to insert $\mathbf{R}_{i+n}^{\frac{1}{2}} \mathbf{R}_{i+n}^{-\frac{1}{2}}$ into the numerator

$$\frac{|\mathbf{w}^H \mathbf{s} \mathbf{v}(\theta)|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} = \frac{\left| s \mathbf{w}^H \mathbf{R}_{i+n}^{\frac{1}{2}} \mathbf{R}_{i+n}^{-\frac{1}{2}} \mathbf{v}(\theta) \right|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} \quad (9.30)$$

Then employ the Schwarz inequality: $|\mathbf{a}\mathbf{b}|^2 \leq |\mathbf{a}|^2 |\mathbf{b}|^2$

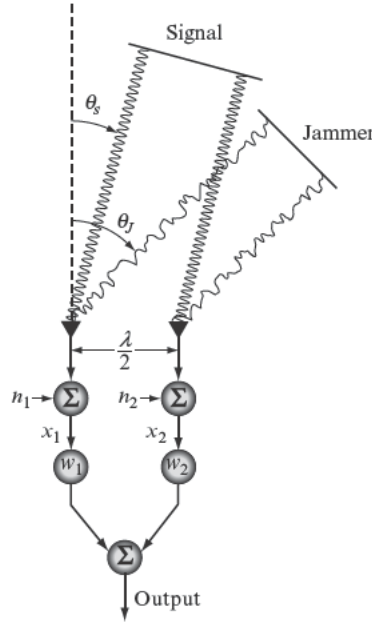
By setting $\mathbf{a} = \mathbf{w}^H \mathbf{R}_{i+n}^{\frac{1}{2}}$ and $\mathbf{b} = \mathbf{R}_{i+n}^{-\frac{1}{2}} \mathbf{v}(\theta)$, the SINR expression can be converted into the inequality

$$\frac{P_s}{P_{i+n}} \leq \frac{|s|^2 \left| \mathbf{w}^H \mathbf{R}_{i+n}^{\frac{1}{2}} \right|^2 \left| \mathbf{R}_{i+n}^{-\frac{1}{2}} \mathbf{v}(\theta) \right|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} = |s|^2 \mathbf{v}^H(\theta) \mathbf{R}_{i+n}^{-1} \mathbf{v}(\theta) \quad (9.31)$$

The weight vector that achieves the bound is

$$\mathbf{w}_{SNR} = k \mathbf{R}_{i+n}^{-1} \mathbf{v}(\theta) \quad (9.32)$$

FIGURE 9-24 ■
Two-weight
example.



which is very similar to the Wiener filter except that the covariance matrix of the interference plus noise is inverted instead of the total covariance matrix.

To see the relationship between these two solutions for the weights, consider a simple two weight example shown in Figure 9-24. Let the signal and jammer signals be

$$s(t) = \sqrt{P_s} e^{j(\omega_0 t + \psi_s)}$$

$$J(t) = \sqrt{P_J} e^{j(\omega_0 t + \psi_J)}$$

and the noise signals, n_1 and n_2 , be independent and identically distributed zero mean Gaussian with variance σ^2 .

The steering vector based on this array geometry is

$$\mathbf{v}(\theta) = [1 \ e^{j\pi \sin \theta}]^H$$

the covariance matrix is

$$\mathbf{R}_x = E[\mathbf{xx}^H] = \begin{bmatrix} P_s + P_J + \sigma^2 & P_s e^{j\pi \sin \theta_s} + P_J e^{j\pi \sin \theta_J} \\ P_s e^{-j\pi \sin \theta_s} + P_J e^{-j\pi \sin \theta_J} & P_s + P_J + \sigma^2 \end{bmatrix}$$

and the inverse of the covariance matrix is

$$\mathbf{R}_x^{-1} = \frac{1}{D_1} \begin{bmatrix} P_s + P_J + \sigma^2 & -P_s e^{j\pi \sin \theta_s} - P_J e^{j\pi \sin \theta_J} \\ -P_s e^{-j\pi \sin \theta_s} - P_J e^{-j\pi \sin \theta_J} & P_s + P_J + \sigma^2 \end{bmatrix}$$

where

$$D_1 = 4P_s P_J \sin^2 \left(\frac{\pi (\sin \theta_s - \sin \theta_J)}{2} \right) + 2P_s \sigma^2 + 2P_J \sigma^2 + \sigma^4$$

and plugging into the equation for the Wiener filter gives

$$\mathbf{w}_{mse} = \mathbf{R}_x^{-1} \mathbf{v}(\theta_s) = \frac{1}{D_1} \begin{bmatrix} \sigma^2 + P_J (1 - e^{j\pi(\sin\theta_J - \sin\theta_s)}) \\ (\sigma^2 + P_J) e^{-j\pi \sin\theta_s} - P_J e^{-j\pi \sin\theta_J} \end{bmatrix}$$

Now let's compare the maximum SINR solution. The inverse of the interference-plus-noise covariance matrix is given by

$$\mathbf{R}_n^{-1} = \frac{1}{D_2} \begin{bmatrix} P_J + \sigma^2 & -P_J e^{j\pi \sin\theta_J} \\ -P_J e^{-j\pi \sin\theta_J} & P_J + \sigma^2 \end{bmatrix}$$

where

$$D_2 = 2P_J\sigma^2 + \sigma^4$$

and plugging this into the equation for the maximum SINR solution gives

$$\mathbf{w}_{SNR} = \mathbf{R}_n^{-1} \mathbf{v}(\theta_s) = \frac{1}{D_2} \begin{bmatrix} \sigma^2 + P_J (1 - e^{j\pi(\sin\theta_J - \sin\theta_s)}) \\ (\sigma^2 + P_J) e^{-j\pi \sin\theta_s} - P_J e^{-j\pi \sin\theta_J} \end{bmatrix}$$

In comparing \mathbf{w}_{mse} and \mathbf{w}_{SNR} , notice that they are identical except for the scaling factors of D_1 and D_2 , respectively. Since scaling factors applied to the weights do not impact the SINR, these two solutions are effectively equivalent.

The value of the maximum SINR can also be computed by

$$\begin{aligned} \text{SINR}_{max} &= P_s \mathbf{v}^H(\theta_s) \mathbf{R}_n^{-1} \mathbf{v}(\theta_s) \\ &= \frac{2\chi_s}{2\chi_J + 1} (1 + \chi_J (1 - \cos(\pi(\sin\theta_J - \sin\theta_s)))) \end{aligned}$$

where

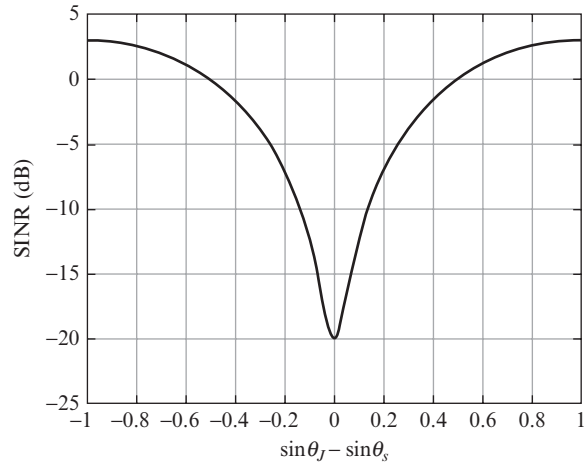
$$\begin{aligned} \chi_s &= \frac{P_s}{\sigma^2} = \text{Input SNR} \\ \chi_J &= \frac{P_J}{\sigma^2} = \text{Input JNR} \end{aligned}$$

Figure 9-25 shows a plot of the SINR as a function of the separation between the target and jammer for in input SNR of 0 dB and an input JNR of 20 dB. As the target–jammer separation goes to zero, the adaptive beamformer is not able to spatially null the jammer, and the output SINR approaches the ratio of the input SNR to the input JNR (χ_s/χ_J).

9.3.3 Constrained Optimization

The Wiener filter and the maximum SINR solutions for the adaptive weights both effectively maximize the SINR for the beam pointed in a particular commanded steer direction, but otherwise they in no way constrain the antenna pattern response. For radar, this can lead to highly distorted antenna patterns in some cases, so often there is a desire to further constrain the antenna pattern to be in some sense good. This additional pattern control can be achieved through using constrained optimization to compute the weights.

FIGURE 9-25 ■
SINR as a function
of target–jammer
separation.



9.3.3.1 Linearly Constrained Minimum Variance

The linear constrained minimum variance (LCMV) solution for the weights minimizes the output power of the beamformer

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_x \mathbf{w} \quad (9.33)$$

subject to a set of constraints defined by

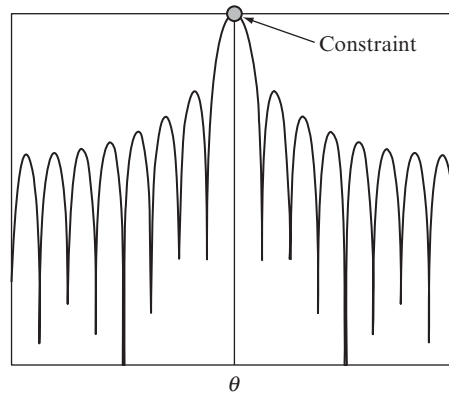
$$\mathbf{C}^H \mathbf{w} = \mathbf{g} \quad (9.34)$$

where the columns of the constraint matrix, \mathbf{C} , are typically steering vectors in the directions of the desired pattern constraints, and the corresponding values in the response vector, \mathbf{g} , are the desired antenna pattern responses in those directions [24]. This constrained optimization problem can be solved using the method of Lagrange multipliers to be

$$\mathbf{w}_{LCMV} = \mathbf{R}_x^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{R}_x^{-1} \mathbf{C})^{-1} \mathbf{g} \quad (9.35)$$

For example, let \mathbf{C} be equal to the steering vector in the commanded steer direction, $\mathbf{v}(\theta_0)$, and \mathbf{g} be equal to 1 for unity gain in that direction as shown in Figure 9-26. This

FIGURE 9-26 ■
Antenna pattern
constraint.



constraint fixes the gain in the center of the beam and then, by minimizing the power at the output of the beamformer, cancels signals coming from any other direction. This special case, where $\mathbf{C} = \mathbf{v}(\theta_0)$ and $\mathbf{g} = 1$, is called the minimum variance distortionless response (MVDR) beamformer. Plugging these values into the LCMV equation results in the following MVDR weight equation:

$$\mathbf{w}_{MVDR} = \frac{R_x^{-1} \mathbf{v}(\theta_0)}{\mathbf{v}^H(\theta_0) R_x^{-1} \mathbf{v}(\theta_0)} \quad (9.36)$$

Notice that the numerator of the equation for the MVDR beamformer is the same as the Wiener filter and the denominator reduces to a scalar. This means that it is also equivalent to the MMSE and maximum SINR solutions to within a scale factor.

Additional constraints can be inserted by adding more columns to the constraint matrix and providing the corresponding pattern response values in \mathbf{g} . Each additional constraint, however, takes away an adaptive DOF that could be otherwise used for cancelling jammers. In the extreme case where the number of constraints is equal to the number of spatial channels, the beamformer becomes data independent.

9.3.3.2 Generalized Sidelobe Canceller

The generalized sidelobe canceller (GSC) is mathematically equivalent to the LCMV but is a useful formulation for analyzing and developing constrained algorithms [25]. In the GSC, the adaptive beamformer is partitioned into an upper data independent branch, \mathbf{w}_f , – and a lower adaptive branch, as shown in Figure 9-27. In the lower branch, a blocking matrix prevents the target signal of interest from entering the adaptive branch. The blocking matrix, \mathbf{B} , is defined as the M by $M-L$ orthogonal complement to the constraint matrix, \mathbf{C} , where M is the number of spatial channels, and L is the number of constraints (i.e., columns in \mathbf{C}). After the blocking matrix, the adaptive weight vector \mathbf{w}_a is unconstrained.

$$\mathbf{w}_a = (\mathbf{B}^H \mathbf{R}_x \mathbf{B})^{-1} \mathbf{B}^H \mathbf{R}_x \mathbf{w}_f \quad (9.37)$$

An example of an adapted pattern from the GSC is shown in Figure 9-28 for a 16-element uniform linear array. In this example, there is a look direction constraint at 10 degrees where there is a 31 dB SNR target and two jammers are located at –40 degrees and –60 degrees. The antenna pattern response with no adaptive beamforming is shown as a green dash-dot line and yields an SINR of –7.7 dB. The GSC response is shown as a blue solid line and has nulls in the jammer directions so that the SINR is increased to 30.8 dB.

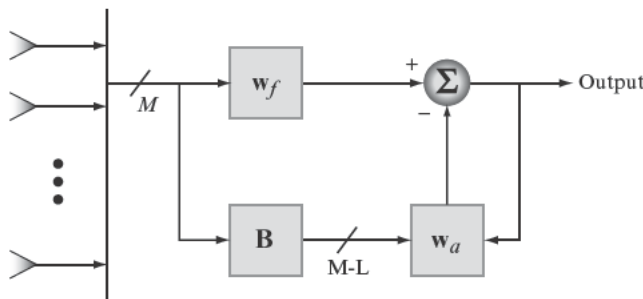
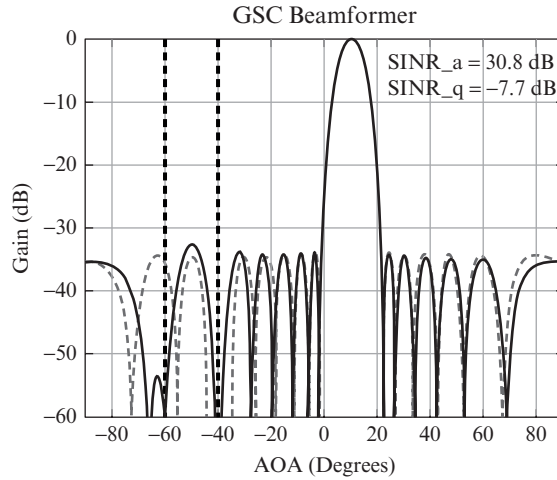


FIGURE 9-27 ■
Generalized
sidelobe canceller.

FIGURE 9-28 ■
Generalized sidelobe
canceller example.



9.3.3.3 Beamformer Constraint Design

In addition to point constraints, other types of constraints are possible to control the antenna pattern response over extended angular regions. Two constraints of this type are derivative constraints and eigenvector constraints.

Derivative Constraints As the name implies, a point constraint will fix the pattern response at a specified point in the antenna pattern (e.g., the center of the mainbeam), but it in no way ensures that there will be a peak at that point. A derivative constraint is a way to force an antenna pattern peak to be located in the constraint direction [26,27]. The derivative constraint is implemented by adding the derivative of the steering vector with respect to angle (evaluated at the constraint angle) as a column in the constraint matrix and adding a zero to the response vector to force a relative maximum in that direction. Note that a point constraint in the look direction is still necessary to maintain gain in that direction.

$$\underbrace{\begin{bmatrix} \mathbf{v}(\theta_0), \frac{\partial \mathbf{v}(\theta_0)}{\partial \theta} \end{bmatrix}^H}_{\mathbf{C}} \underbrace{\mathbf{w}}_{\mathbf{g}} = \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\mathbf{g}}$$

Another application for using derivative constraints is in the difference pattern for monopulse to maintain the monopulse slope at the beam center. In this case, instead of setting the corresponding value of the response vector to zero, it is set equal to the nominal value of the difference pattern slope at beam center. Also, for large arrays, derivative constraints should be used with caution because the derivative of the steering vector can become very large and cause the constraint matrix to lose rank and lead to numerical problems when computing the weights.

Eigenvector Constraints One approach to constraining the antenna pattern over an angular region (as shown in Figure 9-29) might be to place a number of point constraints distributed over the angles in that region. This tends to be a highly inefficient use of degrees of freedom, so a better approach is to employ eigenvector constraints [28]. The constraints are computed by first forming a constraint covariance matrix, which is computed as the

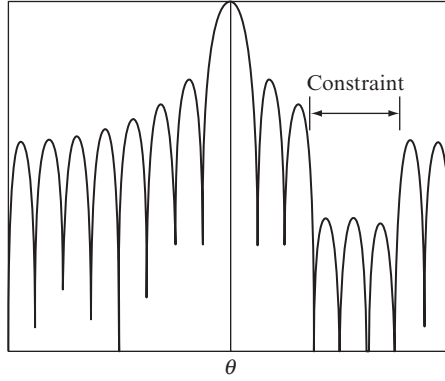


FIGURE 9-29 ■ Pattern constraints over an angular region.

outer product of the steering vectors integrated over the bandwidth and the angular extent of the constraint region.

$$\mathbf{R}_c = \int_f \int_{\theta} \mathbf{v}(\theta, f) \mathbf{v}^H(\theta, f) d\theta df \quad (9.38)$$

By performing an eigenvector decomposition of this covariance matrix, the beamformer constraints can be chosen as the eigenvectors corresponding to the principle eigenvalues of the constraint covariance matrix.

$$\mathbf{C}^H \mathbf{w} = [\mathbf{v}(\theta_0), \mathbf{q}_1, \dots, \mathbf{q}_k]^H \mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $[\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k]$ are the principal eigenvectors of \mathbf{R}_c . The advantage of this approach is that it provides a systematic way of controlling the pattern over an angular region using a minimum number of constraints necessary to achieve the constraint.

9.3.4 Adaptive Weight Estimation

Given an equation for the optimum weights, whether it is the Wiener filter, maximum SINR, or GSC, it is then necessary to estimate the weights with the receive data available in the signal processor. In their standard forms, all the adaptive algorithms involve estimating a covariance matrix, inverting the covariance matrix, and performing some matrix–vector multiplies to produce the adaptive weights. Any errors that cause the estimated adaptive weights to deviate from the optimal weights result in degraded SINR performance due to signal loss or jammer leakage or both.

Since the covariance matrix is not known a priori, it must be estimated from a limited set of data available to train the adaptive weights. If the weights are to be updated from pulse to pulse, then the training data might be all of the data samples (range cells) in the receive window for each pulse or a subset of that data. The details of sample selection for adaptive weight training tend to be very application specific, but key considerations usually include the following:

1. Getting enough training samples for a good covariance matrix estimate
2. Excluding target signal from the training data set

3. Excluding other interfering signals such as clutter if they can be more effectively mitigated in some other way
4. Updating the weights frequently enough to keep up with changes in the jamming environment

The number of samples necessary to get a good estimate of the weights will be discussed in the following section on sample matrix inversion. In theory, the target signal would be prevented from distorting the adaptive weights by the constraint in the commanded beam direction. In practice, the target signal is most likely not located exactly at beam center where the constraint is located, nor is the steering vector used to form that constraint perfectly known due to residual miscalibration in the array. As a result, a strong target signal can leak through the constraint allowing the adaptive algorithm to suppress the target along with the jamming. Target signal cancellation can sometimes be prevented by performing the adaptive beamforming prior to pulse compression. If the uncompressed target is well below the noise, then it won't impact the adaptive weights.

Clutter may be excluded from the adaptive weight training if it would be better handled by Doppler filtering in the subsequent signal processing. This is often the case for surface based radars. Since clutter is typically distributed in angle, it is likely to consume a large number of spatial DOF. If it can be sufficiently mitigated with separate Doppler filtering, then it may be preferable to reserve the limited spatial DOF solely for jamming. A moving target indication filter can be used in the weight training path to provide screen the clutter from the training samples prior to estimating the adaptive weights.

If only a subset of the data is used for adaptive weight estimation, then care must be taken to update the weights frequently enough to keep up with the evolving interference environment. Radar platform motion, jammer platform motion, and blinking jamming can each cause the jamming environment as viewed by the radar to change between the time the weights are computed and the time the weights are applied. Unsurprisingly, allowing the weights to become stale can have a catastrophic impact on the ability of the weights to cancel the jamming.

9.3.4.1 Sample Matrix Inversion

With sample matrix inversion, the estimated covariance matrix is substituted directly into the equation for the optimum weights. The maximum likelihood estimate of the covariance matrix is given by

$$\hat{\mathbf{R}}_x = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{x}_k \mathbf{x}_k^H$$

where the outer product of the data matrix with itself is averaged over some number of data samples K . The number of samples necessary for a good estimate of the weights can be determined by calculating the SINR loss due to estimating the weights relative to the optimum SINR [29]. Figure 9-30 shows the expected value of the SINR loss as a function of the ratio of the number of data samples (snapshots) to the number of DBF channels. For the SINR loss to be no more than 0.5 (or 3 dB) in the mean requires that the number of data samples be around twice the number of spatial channels. In practice, the number of training samples is usually set to 10 or more times the number of spatial channels to get much closer to optimal performance.

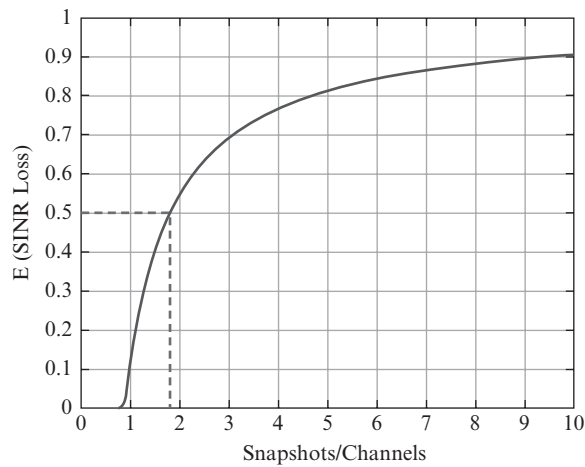


FIGURE 9-30 ■
SINR loss as a
function of the ratio
of data samples to
channels.

9.3.4.2 Weight Jitter Stabilization

In addition to SINR loss, errors in estimating the adaptive weights can produce elevated sidelobes that fluctuate from weight update to weight update. This is known as weight jitter, and it results from poor estimates of the noise eigenvalues in the covariance matrix estimate. The weight jitter can be stabilized by loading the diagonal of the covariance matrix strictly for the purpose of computing the weights [30,31].

$$\tilde{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I} \quad (9.39)$$

Diagonal loading has the effect of swamping out the noise eigenvalues with a fixed synthetic noise level. Figure 9-31 shows the eigenvalue decomposition for a covariance matrix with two different levels of diagonal loading indicated. The desired level would be just high

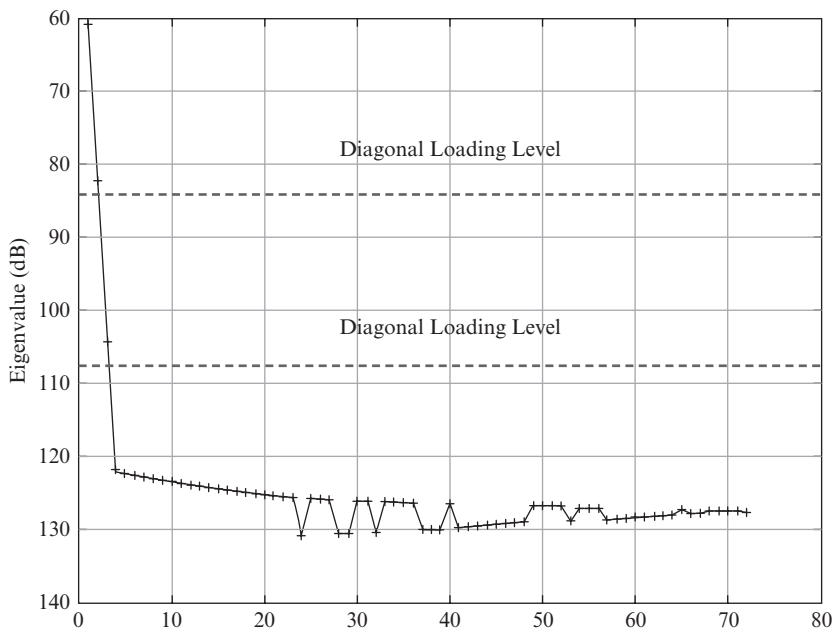


FIGURE 9-31 ■
Diagonal loading.

enough to cover all the noise eigenvalues without eclipsing any eigenvalues due to jammers. This is because the higher the diagonal loading level, the less jammer cancellation is achieved.

9.3.4.3 Cholesky Factorization

The adaptive weight estimation problem can be formulated as solving a system of linear equations.

$$\mathbf{R}_x \mathbf{w} = \mathbf{v}$$

Cholesky factorization is a computationally efficient way of solving this problem in $O(M^3/3)$ operations [32]. The first step is to decompose the covariance matrix as

$$\mathbf{R}_x = \mathbf{L} \mathbf{D} \mathbf{L}^H$$

where \mathbf{L} is a lower triangular matrix, and \mathbf{D} is a diagonal matrix. Software libraries for performing this decomposition are widely available [33]. Using this decomposition of the covariance matrix, the system of equations can be rewritten as

$$\mathbf{L} \mathbf{y} = \mathbf{v}$$

where $\mathbf{y} = \mathbf{D} \mathbf{L}^H \mathbf{w}$. Since \mathbf{L} is a triangular matrix, and \mathbf{v} is the known steering vector, it is easy to solve for \mathbf{y} by applying back substitution. Once \mathbf{y} is known, the weights are computed by solving the second triangular system using back substitution.

$$\mathbf{L}^H \mathbf{w} = \mathbf{D}^{-1} \mathbf{y}$$

The covariance based methods for estimating the adaptive weights have the disadvantage that they involve squaring the data. This squaring causes an increase in dynamic range that can lead to numerical problems. If computing precision is limited, then there are methods using singular value decomposition or QR decomposition that solve for the adaptive weights directly from the data without squaring it [34].

9.3.5 Performance Metrics

Performance metrics are used to assess the effectiveness of the adaptive filter in cancelling a given jamming scenario. This section will define the most common metrics used for evaluating adaptive cancellation performance. Unfortunately, some of these metrics can be precisely computed only in simulation and can be estimated or inferred only from hardware measurements. Most of the performance metrics involve taking the ratio of various combinations of the signal, jammer, and noise powers with and without adaptive cancellation. Let S_u , J_u , and N_u be the unadapted signal, jammer, and noise powers, respectively, and S_a , J_a , and N_a be the adapted signal, jammer, and noise powers, respectively:

- N_u = unadapted noise only
- $S_u + N_u$ = unadapted signal plus noise
- $J_u + N_u$ = unadapted jammer plus noise
- $J_a + N_a$ = adapted jammer plus noise
- $S_a + J_a + N_a$ = adapted signal plus jammer plus noise

9.3.5.1 Cancellation Ratio

The cancellation ratio (CR) is defined as the ratio of the jammer power without adaptive cancellation to the jammer power with cancellation

$$CR_J = \frac{J_u}{J_a} = \frac{\mathbf{w}_0^H R_J \mathbf{w}_0}{\mathbf{w}_a^H R_J \mathbf{w}_a} \quad (9.40)$$

Cancellation ratio provides a measure of how well the jamming is suppressed by adaptive cancellation. Typically, CR is used as a metric to evaluate how well the hardware in the receive chain is calibrated and matched to support the desired level of cancellation. Unfortunately, CR as previously defined cannot be measured directly on real hardware implementations because the jammer cannot be measured without the noise also being present.

An alternative form of the CR that can be directly measured on hardware is

$$CR_{J+N} = \frac{J_u + N_u}{J_a + N_a} \quad (9.41)$$

This version of the cancellation ratio converges to the jammer-only version when $J_u, J_a \gg N_u, N_a$, but it puts a ceiling on the cancellation ratio as J_a becomes very small. The jammer-only cancellation ratio can be inferred from the jammer-plus-noise cancellation ratio if the adaptive weights that are computed when the jammer is turn on are stored and then applied to noise-only data when the jammer is turned off to estimate the adapted noise power (N_a). Then CR_J can be estimated by subtracting N_u and N_a from the numerator and denominator of CR_{J+N} , respectively. This approach works under the conditions that the jammer and noise signals are uncorrelated and that the noise power is stationary over the time the measurements are taken.

9.3.5.2 Residue

The residue is usually defined as the output power of the beamformer divided by the noise power and typically does not include target signal

$$\text{Residue} = \frac{J + N}{N} \quad (9.42)$$

The residue can be computed with and without adaption to show how well the jammer is being suppressed. If the jammer is perfectly cancelled, then the residue will be 0 dB. The residue is useful for comparing adaptive cancellation approaches based on the amount of jamming that remains after cancellation.

9.3.5.3 SINR

SINR is the ratio of the signal power to the sum of the jammer and noise powers.

$$\text{SINR} = \frac{S}{J + N} \quad (9.43)$$

This provides a better system-level measure of performance because it also captures the impact of the adaptive beamformer on the signal of interest. Cancellation ratio and residue measure only what happens to the jammer, so if the adaptive beamformer suppresses the signal along with the jammer that will not be revealed by those metrics. SINR, on the other

hand, includes the effects on the signal. SINR is also useful because it is directly related to the probability of detection and the target parameter estimation error.

To gauge how much benefit the adaptive canceller provides relative to no adaption, the SINR improvement can be calculated as

$$\text{SINR}_{imp} = \frac{\text{SINR}_a}{\text{SINR}_u} \quad (9.44)$$

Alternatively, to see how close the adaptive beamformer comes to achieving optimal performance, it is useful to calculate the SINR loss

$$\text{SINR}_{loss} = \frac{\text{SINR}_a}{\text{SNR}} = \underbrace{\frac{\text{SINR}_{opt}}{\text{SNR}}}_{L_{s,1}} \cdot \underbrace{\frac{\text{SINR}_a}{\text{SINR}_{opt}}}_{L_{s,2}} \quad (9.45)$$

The SINR loss can be factored into two terms: the clairvoyant SINR loss ($L_{s,1}$); and the adaptive SINR loss ($L_{s,2}$). The clairvoyant SINR loss compares the interference-limited performance to the noise-limited performance. The adaptive SINR loss compares the adaptive performance based on estimated statistics to the optimum performance based on known statistics.

9.3.5.4 Antenna Patterns

Antenna patterns also reveal important aspects of the adaptive filter performance including the following:

- Null depth in the jammer location
- Gain in the target direction
- Post-adapted sidelobe levels

Individual antenna patterns are useful for spot-checking specific jamming scenarios but are cumbersome for assessing the global performance over all possible jammer scenarios. Also, when the adaptive beamforming is at the subarray level, there may be sweet and sour spots in performance depending on where the jamming is located in angle. Therefore, individual antenna patterns may give either an overly optimistic or pessimistic view of the overall cancellation performance depending on where the jamming is located in angle.

9.3.5.5 Cumulative Distribution Functions

To get a more global perspective of performance, it is necessary to perform Monte Carlo simulation where the jammer locations are randomized over the span of possible angles of arrival. Then the cumulative distribution function of performance metrics such as SINR can be calculated to show the probability that certain performance thresholds are met over the ensemble of possible interference scenarios. Figure 9-32 shows an example of the cumulative distribution function curves of SINR for both an unadapted and an adapted beamformer over a set of Monte Carlo trials where a single jammer was randomly located in the sidelobes of the antenna pattern. For this example, the noise-limited SNR was 22 dB, so in the plot a vertical line at 22 dB would indicate perfect cancellation. The curves show the probability that the SINR is less than a particular value. For example, if there were a requirement that the SINR be greater than 15 dB, then from the plot the probability that the SINR is less than 15 dB is 75% without adaptive cancellation but is less than 10% with

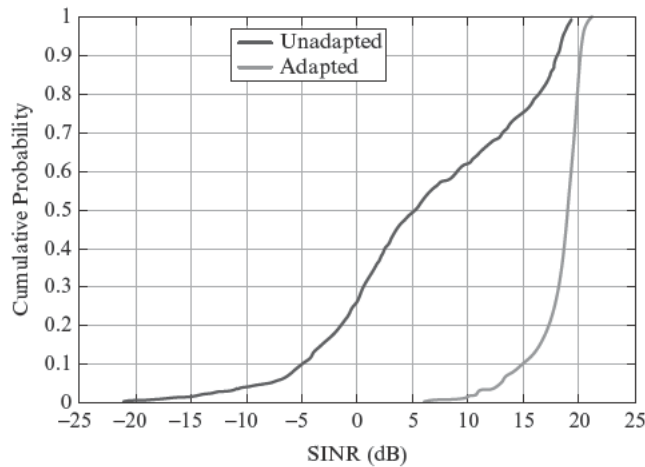


FIGURE 9-32 ■
Example cumulative
distribution function
of SINR.

adaptive cancellation. This way of displaying the performance metrics from Monte Carlo trials is a convenient way to compare different algorithm options or subarray designs or to assess the impact of various types of errors or hardware limitations.

9.4 | ADAPTIVE BEAMFORMER ARCHITECTURES

An adaptive beamformer weights and combines spatially distributed channels to form beams, but the configuration or architecture of those spatial channels has a large impact on how the adaptive beamformer performs. In this section, the fundamental beamformer architectures are discussed along with their relative advantages and disadvantages.

9.4.1 Sidelobe Blanker

The sidelobe blanker is not actually an adaptive cancellation technique; however, it has a lot of similarities with the sidelobe canceller described in the following section, so it is helpful to understand what a sidelobe blanker does and how it is different from adaptive cancellation before discussing the true adaptive beamformer architectures. Figure 9-33 shows a schematic diagram of the sidelobe blanker.

The sidelobe blanker is implemented by using one or more low gain auxiliary antennas to confirm that a signal giving rise to a detection is in the mainbeam and not a strong interference source or large target in the sidelobe region. This is done by comparing the magnitude of the main antenna output to the magnitude of the auxiliary-blanker antenna

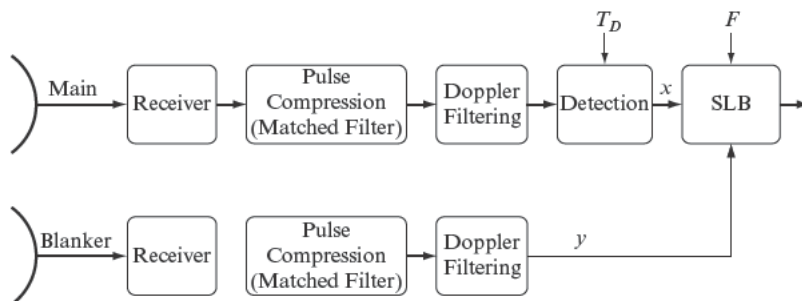
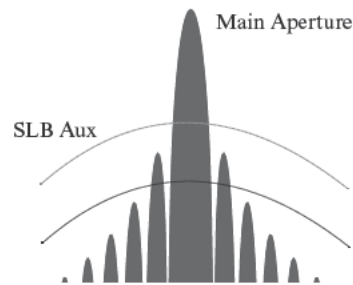


FIGURE 9-33 ■
Sidelobe blanker
schematic diagram.

FIGURE 9-34 ■
Notional antenna
patterns for a
sidelobe blanker.



output. If the output of the main channel is larger, that should occur only when the signal is in the mainbeam because the gain in the main antenna is greater than the gain in the auxiliary antenna. However, if the output of the auxiliary antenna is greater than the output of the main antenna, that should happen only when the signal originates from somewhere in the sidelobes. Figure 9-34 shows a notional main aperture antenna pattern along with two choices for the sidelobe blanker auxiliary pattern. For the lower gain auxiliary pattern, the first sidelobes of the main aperture exceed the auxiliary pattern and that will cause the sidelobe blanker to fail for interference in those regions resulting in a higher probability of false detections. Adding gain to the auxiliary channel can eliminate these “stick-throughs” as shown in the alternative auxiliary pattern, but that also increases the noise in the auxiliary channel resulting in a higher probability of false blanking.

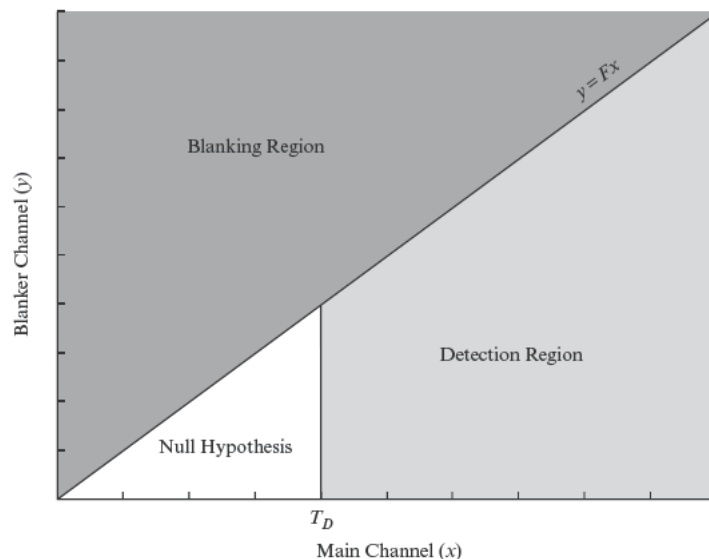
A simple way to construct this hypothesis test is to compare the ratio of the blanker output, y , to the main channel output, x , to a threshold value, say, F .

$$\frac{y}{x} > F : \text{Blank the signal}$$

Otherwise : Process the detection

Figure 9-35 shows the blanking and detection regions for the sidelobe blanker based on this threshold test. The sidelobe blanker provides a way to flag and excise returns that

FIGURE 9-35 ■
Blanking and
detection regions for
sidelobe blanker.



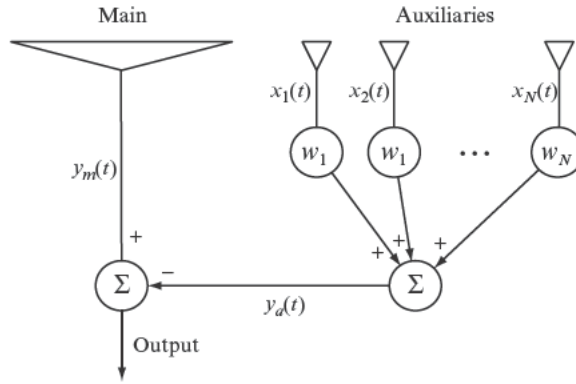


FIGURE 9-36 ■
Sidelobe canceller
architecture

have been corrupted by sidelobe signals (jamming or large targets). It does not cancel the jamming in the way that the adaptive beamforming methods do, which allows that data to continue to be processed for mainbeam detections and parameter estimation. Often the sidelobe blanker will be used in conjunction with adaptive beamforming to remove sidelobe coherent repeater jammers or intermittent interference that the adaptive beamforming is unable to cancel effectively. In this case, it is necessary to adaptively null the blanker pattern in addition to the sum receive beam so that jamming that is being effectively canceled does not trigger the blanker. For more information on the sidelobe blanker see [35].

9.4.2 Sidelobe Canceller

The sidelobe canceller was originally developed in the 1950s and was the first adaptive cancellation technique to be applied to radar [36]. It consists of a main channel antenna and a number of low gain auxiliary antennas that are spatially distributed around the main channel, as shown in Figure 9-36. The main channel can be any type of antenna including a reflector, a conventional phased array, or a DBF array. If the main channel is a phased array, then the auxiliary channels may be individual elements or small groups of elements that are tapped off separately and connected to dedicated receivers. The auxiliary antennas are adaptively weighted so that the resulting combined auxiliary signal matches the jamming in the main channel. Therefore, when the combined auxiliary signal is subtracted from the main channel, the jamming is cancelled. The optimal solution for the auxiliary weight vector is given by

$$\mathbf{w}_{aux} = \mathbf{R}_{aux}^{-1} \mathbf{r}_{ma} \quad (9.46)$$

where,

$$\mathbf{R}_{aux} = E[\mathbf{x}_{aux} \mathbf{x}_{aux}^H] \quad (9.47)$$

is the covariance matrix of the auxiliary channels, and

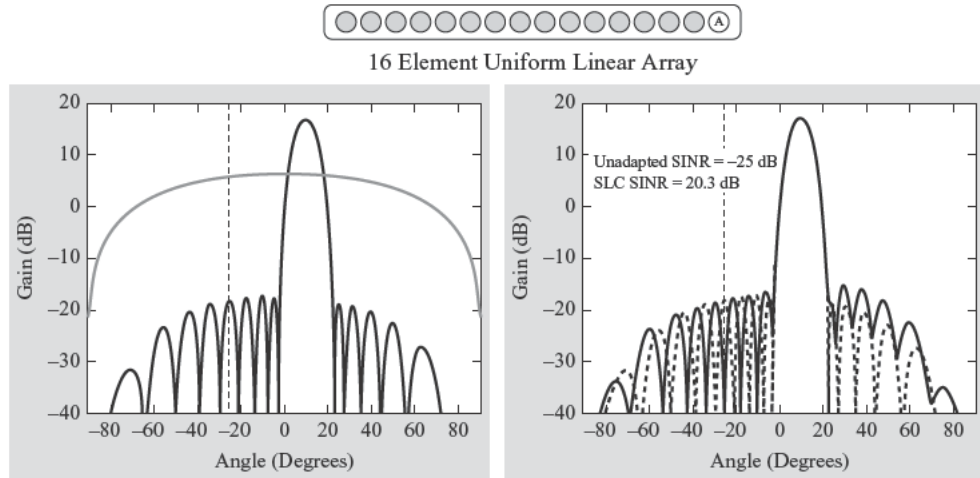
$$\mathbf{r}_{ma} = E[\mathbf{y}_m^* \mathbf{x}_{aux}] \quad (9.48)$$

is the cross-correlation vector between the main channel and the auxiliary channels.

The performance of the sidelobe canceller strongly depends on the number of auxiliaries, their spatial distribution, and the auxiliary antenna pattern. The following examples using a 16-element uniform linear array will be used to illustrate how the auxiliary

FIGURE 9-37 ■

Sidelobe canceller with a single auxiliary. Left plot: dark solid line, main; light solid line, aux; dashed line, jammer. Right plot: dark solid line, adapted pattern; light solid line, weighted aux; dotted line, main; dashed line, jammer.

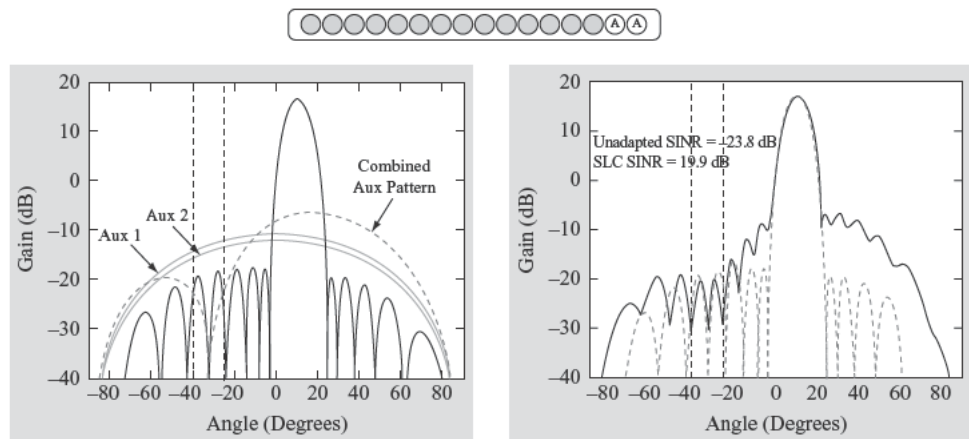


configuration impacts cancellation performance. Figure 9-37 shows a sidelobe canceller example where 15 of the 16 elements in the uniform linear array are being used as the main channel and the remaining edge element is being used as the auxiliary. The plot on the left shows the unadapted patterns for the main and auxiliary channels along with the location of a single jammer indicated by the vertical dashed line. Looking at the sidelobe canceller from the perspective of the antenna patterns, applying the adaptive weight to the auxiliary patterns causes the weighted pattern to match the main channel pattern at the jammer location so that when it is subtracted off there is a null in the jammer direction. The plot on the right shows the post-adapted pattern with a null at the jammer location. Without the sidelobe canceller, the unadapted array has an SINR of -25 dB, but with the SLC the SINR increases to $+20.3$ dB, providing more than 45 dB of SINR improvement.

In the next example, the number of auxiliaries and the number of jammers are increased to two. In this case the two auxiliaries are adjacent elements at the edge of the array. The plot on the left in Figure 9-38 shows the unadapted main pattern, the two auxiliary patterns, and the combined auxiliary pattern after adaptive weighting. The plot on the right shows that the resulting adapted antenna pattern has nulls in the jammer locations and an SINR improvement of nearly 44 dB, but the pattern is degraded at angles away from the jammer

FIGURE 9-38 ■

Sidelobe canceller with two closely spaced auxiliaries. Left plot: dark solid line, main; dashed lines, jammers. Right plot: solid line, adapted pattern; dotted line, main; dashed lines, jammers.



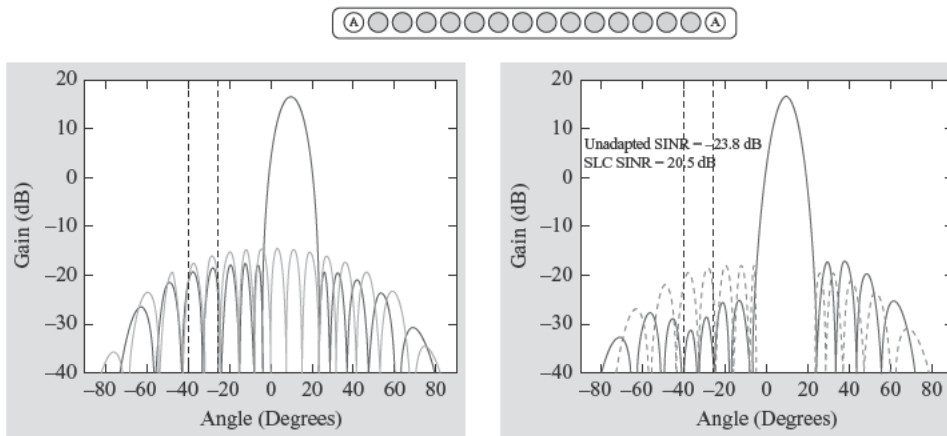


FIGURE 9-39 ■ Sidelobe canceller with two widely space auxiliaries. Left plot: dark solid line, main; light solid line, combined aux; dashed lines, jammers. Right plot: solid line, adapted pattern; dotted line, main; dashed lines, jammers.

locations. The cause of this pattern distortion is apparent in the combined auxiliary pattern shown in the plot on the left. The individual auxiliary patterns, the auxiliary spacing, and the adaptive weights determine the shape of the combined auxiliary pattern. For the relatively closely space jammers in this example, the closely space auxiliaries don't provide enough spatial resolution to cancel both jammers without distorting the pattern at other angles.

The next example improves the adapted pattern performance by moving the auxiliaries to the extreme edges of the array. Increasing the separation between the auxiliaries to be much greater than a half-wavelength causes the combined auxiliary pattern to have a grating structure as shown in the plot on the left in Figure 9-39. This grating structure in the combined auxiliary pattern gives the sidelobe canceller better spatial resolution to null closely spaced jammers with much less degradation of the adapted pattern at other angles. The plot on the right in Figure 9-39 shows that the widely space auxiliaries provide good jammer cancellation (0.6 dB more SINR than the closely spaced auxiliaries), and the adapted pattern is improved.

The performance of the sidelobe canceller is also a function of the main channel sidelobe level in the direction of the jammers. Figure 9-40 shows the jammer residue as a

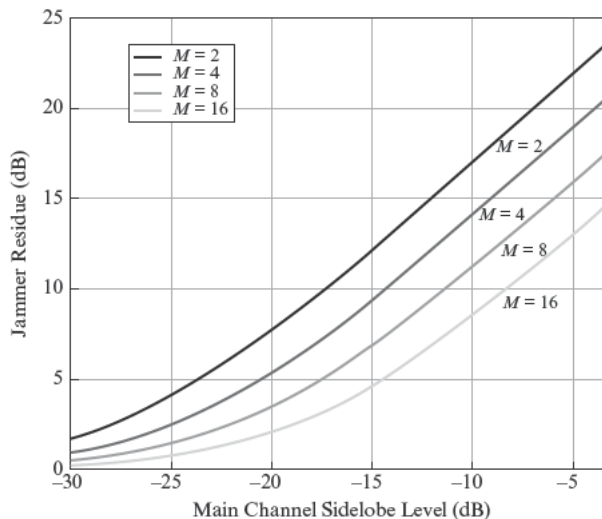
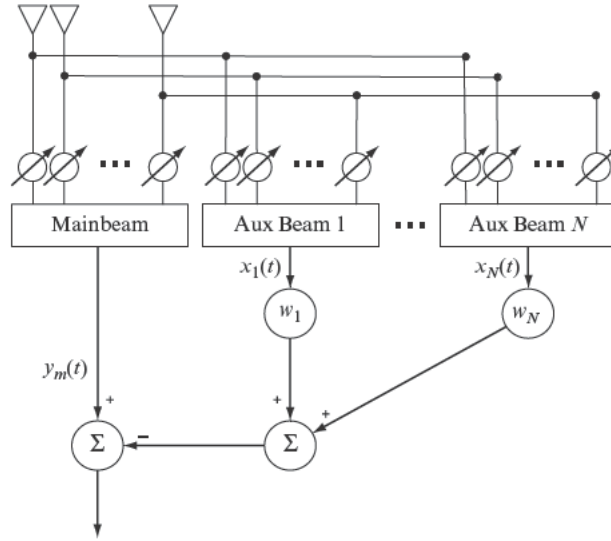


FIGURE 9-40 ■ Sidelobe canceller jammer residue as a function of the main channel sidelobe level at the jammer location for $M = 2$, 4, 8, and 16 single-element auxiliaries.

FIGURE 9-41 ■
Beamspace
adaptive canceller
architecture.



function of the main channel sidelobe level in the direction of a single jammer for sidelobe cancellers with 2, 4, 8, and 16 single-element auxiliaries. As the main channel sidelobe level increases, the jammer residue increases because more gain must be added to the auxiliaries to create the null on the higher sidelobe. The additional gain causes the noise in the auxiliary channels to feed through and increase the residue. Increasing the number of auxiliaries improves performance because that allows multiple auxiliaries to be adaptively combined to cancel on the higher sidelobes.

9.4.3 Beamspace Adaptive Cancellation

The beamspace adaptive canceller is appropriate for radars that use multiple beams on receive. A block diagram of the beamspace adaptive canceller is shown in Figure 9-41. The beamspace canceller is similar to the sidelobe canceller except that it applies adaptive weights to fully formed beams instead of low gain auxiliary antennas. The auxiliary beams are typically steered to fixed positions relative to the mainbeam prior to adaptive processing.

The adaptive weights for the beamspace canceller are given by

$$\mathbf{w}_{beam} = \mathbf{R}_{beam}^{-1} \mathbf{r}_{mb} \quad (9.49)$$

where

$$\mathbf{R}_{beam} = E[\mathbf{x}_{beam} \mathbf{x}_{beam}^H] \quad (9.50)$$

is the covariance matrix of the auxiliary beam outputs, and

$$\mathbf{r}_{mb} = E[y_m^* \mathbf{x}_{beam}] \quad (9.51)$$

is the cross-correlation vector between the mainbeam and the auxiliary beams. Since the auxiliary beams are arranged in a cluster around the mainbeam location, the beamspace canceller is most effective against jamming in or near the mainbeam where the auxiliary beams have the most gain. It is not as effective against jamming in the sidelobes because the jammer will also be in the sidelobes of the auxiliary beams.

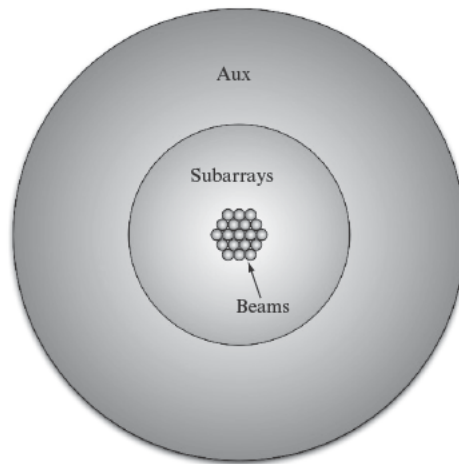


FIGURE 9-42 ■
Regions where each type of DOF provides the best jammer cancellation performance.

9.4.4 Subarray Space Adaptive Cancellation

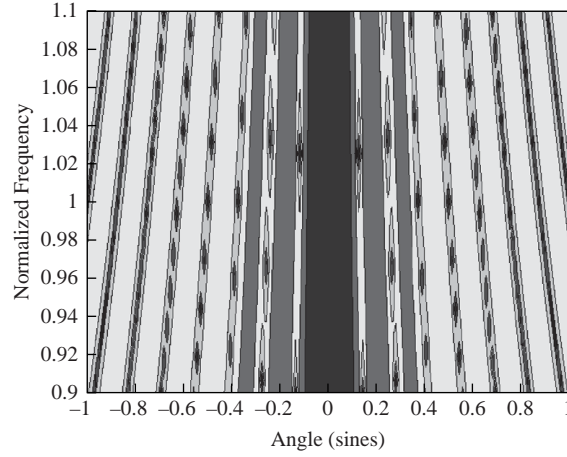
In a subarray space adaptive canceller, the adaptive weights are estimated from and applied to the subarray data. Compared to the subarray space and beam space cancellers, the sidelobe canceller performs well against sidelobe jamming but performs poorly against mainlobe jamming because the auxiliary antennas used for cancellation have very little directivity. Compared to the sidelobe canceller and the subarray space canceller, the beamspace canceller performs well against jamming in or near the mainlobe but performs poorly against sidelobe jammers that are in the sidelobes of all the auxiliary beams. Subarray space is a compromise between those two extremes. It provides the best cancellation performance in the annular region where the subarray pattern exceeds the auxiliary pattern but is lower than the beams used in the beamspace canceller. Figure 9-42 illustrates the regions in angle space where each of these types of DOF (i.e., auxiliaries, subarrays, and beams) provides the best jammer cancellation.

All the adaptive algorithms discussed previously (e.g., Wiener filter, LCMV, GSC) can be applied at the subarray level with the caveat that grating nulls can be a problem due to the greater than half-wavelength spacing of the subarray phase centers. This implies that there will be periodic regions in angle space that are spatially ambiguous with the mainbeam. If a jammer is located in one of these ambiguous regions, then the adaptive weights that put a null on that jammer will also put a null in the mainbeam, causing significant beam distortion and signal loss. To a certain extent, making the subarrays irregular can break up these grating nulls (just like grating lobes), but practical limits on the degree of irregularity that can be imposed on the subarray shapes and sizes usually prevents these undesirable pattern artifacts from being completely eliminated. Overlapped subarrays do not mitigate the effects of grating nulls. For more information on subarray design issues see [14-21].

9.5 | WIDEBAND CANCELLATION

Wideband operation introduces an additional set of challenges for adaptive beamforming. For simplicity, the jammer bandwidth will be assumed to the waveform bandwidth

FIGURE 9-43 ■
Wideband antenna
pattern response.



throughout this section. Over wider bandwidths, the difference in the propagation delay between the digital channels leads to jammer dispersion and frequency response mismatches limit the depth of nulls on the jammers. The frequency response mismatches arise from both the frequency response variations of the hardware components in the receive chain and antenna pattern variations over frequency.

The steering vector is a function of frequency as well as angle:

$$\mathbf{v}(\theta, f) = \left[1 \ e^{j2\pi f \left(\frac{d \sin \theta}{c}\right)} \ \dots \ e^{j2\pi f (K-1) \left(\frac{d \sin \theta}{c}\right)} \right] \quad (9.52)$$

and therefore the antenna pattern response will vary as a function of frequency.

$$F(\theta, f) = \mathbf{w}^H \mathbf{v}(\theta, f) \quad (9.53)$$

A contour plot of a wideband antenna pattern response as a function of angle and normalized frequency is shown in Figure 9-43. As the electrical size of the antenna changes with frequency, the locations of the sidelobe peaks and nulls move in angle. As a result, to cancel jammers over a wide bandwidth, it is necessary to jointly optimize the pattern response in both angle and frequency.

Several possible approaches to achieving cancellation over a wide bandwidth are illustrated in Figure 9-44. The figure represents a region in the sidelobes of the antenna pattern with angle varying along the horizontal axis and frequency along the vertical axis. The shaded areas indicate the null width for the various wideband cancellation options. The dark diagonal line shows the null that would be produced if a narrowband beamformer were applied with weights computed from data only at the center frequency. In this case, the null will be in the correct angular location only at the center frequency and will sweep away from the jammer location for other frequencies across the band. This would allow nearly all the jammer power to leak through the adaptive filter except for a very narrow band around the center frequency. The broad gray region shows the null that would be produced by a narrowband (spatial only) beamformer with weights computed using the full wideband data. In this case, the beamformer uses multiple spatial degrees of freedom to form a broad null that covers the jammer position over the full band.

The light gray regions show the nulls produced if the data are filtered into subbands with separate adaptive beamformers in each subband to cancel the jammer. This effectively

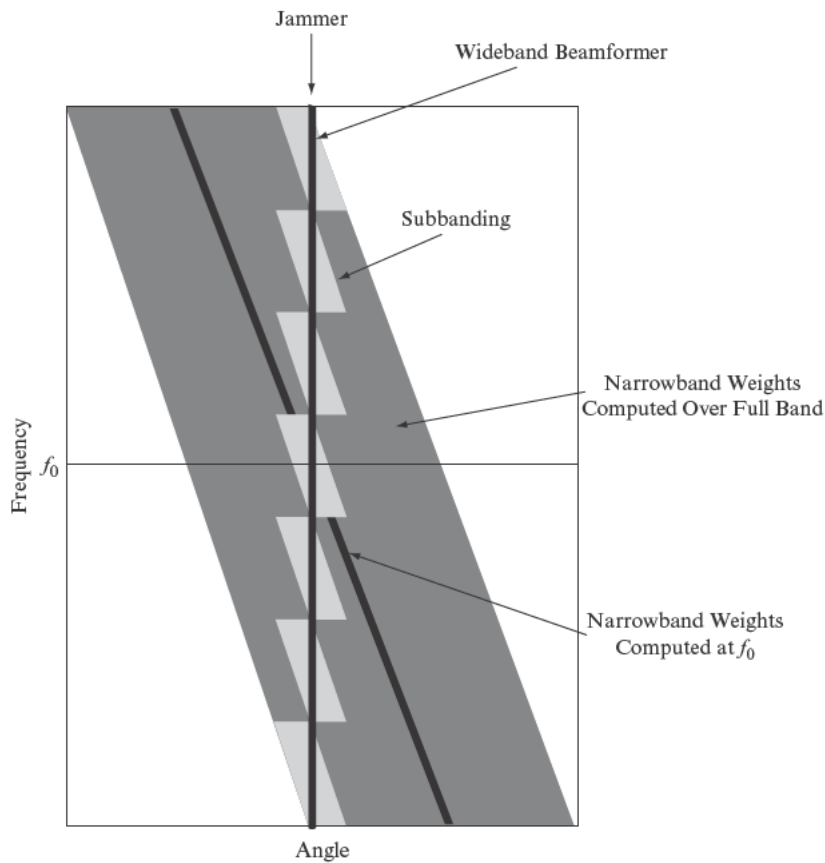


FIGURE 9-44 ■
Nulling regions
associated with
various wideband
cancellation
approaches.

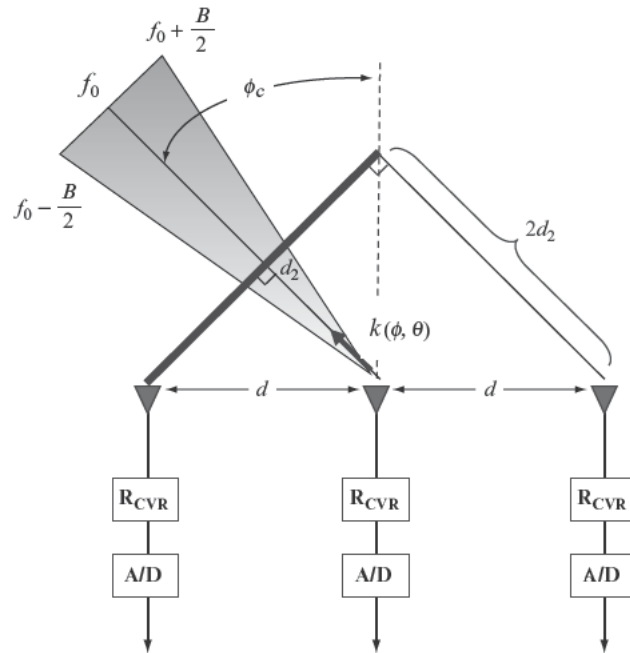
breaks the wideband problem into multiple narrowband cancellation problems. The dark vertical line along the jammer angle shows a joint spatial-frequency wideband adaptive filter that is able to place a sharp null on the jammer location over the entire band.

9.5.1 Jammer Dispersion over Bandwidth

A wideband phased array will be time-delay steered in the direction of the target. The location of the jammer, however, is not known a priori and is presumed not to be colocated with the target if adaptive spatial cancellation is to be employed. Therefore, preset time-delay steering cannot be used to mitigate the jammer's dispersive effects. The dispersive nature of a wideband barrage noise jammer makes it appear as if it is multiple narrowband jammers spread spatially about the jammer's true location. For a wideband jammer, the array response smears across an angular extent proportional to the waveform bandwidth, as depicted in Figure 9-45.

A consequence of spatial spread of the jammer signal is that additional degrees of freedom are required to maintain maximum SINR. Temporal degrees of freedom may be introduced by adding adaptive transversal filters to each channel. The transversal filters offer both additional degrees of freedom and a means of exploiting adaptive time delay to jointly control the frequency response. Alternatively, subbanding allows the limited spatial degrees of freedom to be reused in each subband to compensate for the variation with frequency.

FIGURE 9-45 ■
Wideband jammer
dispersion.



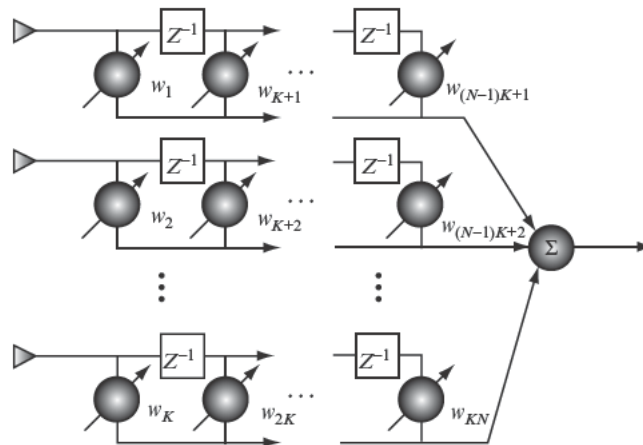
9.5.2 Joint Spatial-Frequency Domain Wideband Beamformer

A spatial adaptive beamformer can be augmented to support wide bandwidths by replacing the single adaptive weight in each spatial channel with a transversal filter, as shown in Figure 9-46. The transversal filters provide additional temporal degrees of freedom needed to jointly control the spatial-frequency response of the array. The required number of taps in the transversal filters is a function of the desired SINR performance, the spatial extent of the array, number of jammers, and the proximity of the jammer to the target. The spacing between taps is usually equal to the ADC sampling rate, T_s .

Like the spatial-only beamformer, the output of this space-time beamformer can be written as

$$\mathbf{y}(t) = \mathbf{w}^H \mathbf{x} \quad (9.54)$$

FIGURE 9-46 ■
Wideband adaptive
beamformer.



where the weight and array data vectors are constructed by concatenating the weights and array data associated with each K spatial channels and N taps resulting in

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \\ w_{K+1} \\ \vdots \\ w_{NK} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_K(t) \\ x_1(t - T_s) \\ \vdots \\ x_K(t - (N-1)T_s) \end{bmatrix} \quad (9.55)$$

If the weights and array data vectors are defined in this manner, then the steering vector is also augmented to be

$$\mathbf{v}(\theta, f) = \begin{bmatrix} 1 \\ e^{-j2\pi f T_s} \\ \vdots \\ e^{-j2\pi f (N-1)T_s} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ e^{j2\pi f (\frac{d \sin \theta}{c})} \\ \vdots \\ e^{j2\pi f (K-1)(\frac{d \sin \theta}{c})} \end{bmatrix} \quad (9.56)$$

where

$$\begin{bmatrix} 1 & e^{-j2\pi f T_s} & \dots & e^{-j2\pi f (N-1)T_s} \end{bmatrix}^T$$

is the temporal steering vector of the transversal filters

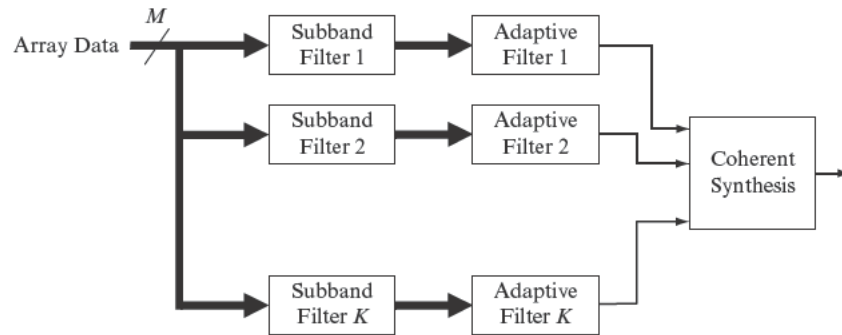
$$\begin{bmatrix} 1 & e^{j2\pi f (\frac{d \sin \theta}{c})} & \dots & e^{j2\pi f (K-1)(\frac{d \sin \theta}{c})} \end{bmatrix}^T$$

is the spatial steering vector and \otimes is the Kronecker product. Once these augmented steering, data, and weight vectors have been formed, they can be inserted directly into all the adaptive algorithms previously discussed in Section 9.3.

All of these narrowband adaptive algorithms are designed to maximize SINR at a single frequency based on the steering vector used to point the beam. For the constrained optimization techniques such as LCMV and the GSC, the steering vector is used to form the point constraint in the beam direction. Given this single point constraint, however, there is nothing to prevent the adaptive algorithm from trading away bandwidth to achieve even marginal gains in SINR. Since the bandwidth is typically selected to provide a required amount of range resolution, this is not usually a desirable trade-off. This is particularly true when the amount of SINR gained is negligible. As the signal bandwidth increases, the adaptive weights in the transversal filter are likely to maximize SINR by forming a narrowband filter response in each subarray channel that attenuates both the signal and jammer bandwidth. The result is both a loss in range resolution and a perturbation of the range sidelobe response. Fortunately, the constrained optimization algorithms provide a convenient means to force the adaptive beamformer to preserve the bandwidth on or near the beam center. This is done by inserting additional constraints into the constraint matrix that are located at the beam center in angle and distributed in frequency across the band. The resulting constraint matrix is given by

$$\mathbf{C} = [\mathbf{v}(\theta_0, f_{min}), \dots, \mathbf{v}(\theta_0, f_0), \dots, \mathbf{v}(\theta_0, f_{max})]$$

FIGURE 9-47 ■
Subband canceller.



and the corresponding response vector, \mathbf{g} , is equal to a vector of ones sized for the number of columns in \mathbf{C} . The number of constraints necessary to preserve the bandwidth increases with the number of taps, and as a rule-of-thumb can be set equal to the number of taps.

The additional DOF provided by the transversal filters significantly enhance the wideband jammer cancellation performance, but at a cost of greatly increasing the number of adaptive weights that need to be estimated. The covariance matrix is inflated to dimension $(NK) \times (NK)$, which increases the computational complexity of the matrix inverse by a factor of the number of taps cubed compared with the spatial-only case. Also, the larger covariance matrix requires proportionally more training samples to maintain equivalent SINR loss due to estimation error.

9.5.3 Subbanding

An alternative to using transversal filters for wideband cancellation is to decompose the spatial channels into a number of subbands and then perform adaptive beamforming separately in each subband [37]. If the subbands can be made sufficiently narrow, then narrowband adaptive beamforming can be used in each subband. Figure 9-47 shows a simplified block diagram of the functional processing flow in a subband canceller. This is equivalent to breaking the wideband problem into many narrowband cancellation problems that are more easily solved and then coherently recombining to provide piecewise wideband cancellation. The subband canceller has a computational advantage compared to the transversal filter approach because it is solving a narrowband problem K times instead of jointly solving the larger wideband problem. This scales the number of operations by a factor of the ratio of the number of subbands to the number of taps cubed. A computationally efficient way to perform the subbanding is to use polyphase filtering [38,39].

9.5.4 LFM/Stretch

For wideband radars that use stretch processing to compress linear frequency modulated (LFM) waveforms, a convenient coupling between time and frequency can be exploited in the design of the adaptive beamformer. Stretch processing consists of an active correlation in which the return signal is demodulated with a replica of the transmit waveform. The response of this deramp operation to a target return signal is a tone at an intermediate frequency proportional to the target range as shown in Figure 9-48. The deramp is followed by a low-pass filter matched to the intermediate frequency bandwidth (B_{IF}) and then an ADC converter as shown in Figure 9-49. For a wideband noise jammer, stretch processing is equivalent to sweeping the low-pass filter along the LFM ramp as illustrated in Figure 9-50.

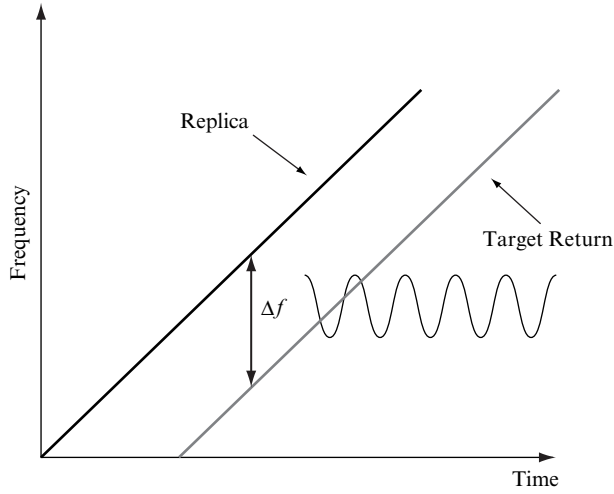


FIGURE 9-48 ■
Deramp of the target return.

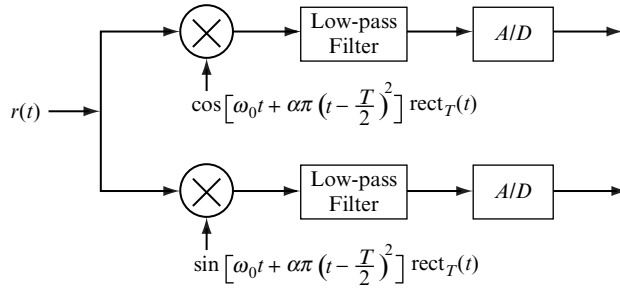


FIGURE 9-49 ■
Stretch processing
block diagram.

In each receiver channel, the jammer signal is modulated by the frequency response of the subarray antenna pattern windowed by the frequency response of the low-pass filter as it sweeps along the LFM ramp. Since in general each subarray pattern is different, time-varying channel-to-channel mismatches are induced on the jammer signal. The time-frequency coupling inherent in stretch processing causes the bandwidth at any instant in time to be relatively narrow.

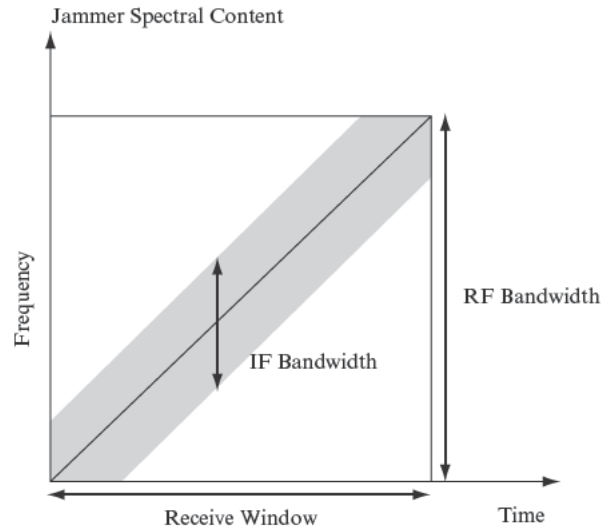
The instantaneous cross-correlation between the m -th and n -th elements in an array for a white noise jammer that has been stretch processed is given by

$$r_{mn}(t) = P_J \text{sinc}\left(\frac{B_{IF}\tau_{mn}}{2}\right) e^{j(\omega_0 + \alpha t)\tau_{mn}} \quad (9.57)$$

where P_J is the jammer power, τ_{mn} is the propagation delay between the two elements, and α is the LFM slope. This cross-correlation is nonstationary due to the time-frequency coupling introduced by mixing the jammer signal with the replica. Averaging this instantaneous correlation over K samples, as would be done when estimating the covariance matrix, gives

$$\bar{r}_{mn} = \frac{1}{K} \sum_{k=0}^{K-1} r_{mn}(t)|_{t=kT_s} = \frac{P_J}{K} \text{sinc}\left(\frac{B_{IF}\tau_{mn}}{2}\right) \frac{\sin\left(\frac{\alpha\tau_{mn}KT_s}{2}\right)}{\sin\left(\frac{\alpha\tau_{mn}T_s}{2}\right)} e^{j\frac{\alpha}{2}(K-1)T_s\tau_{mn}} \quad (9.58)$$

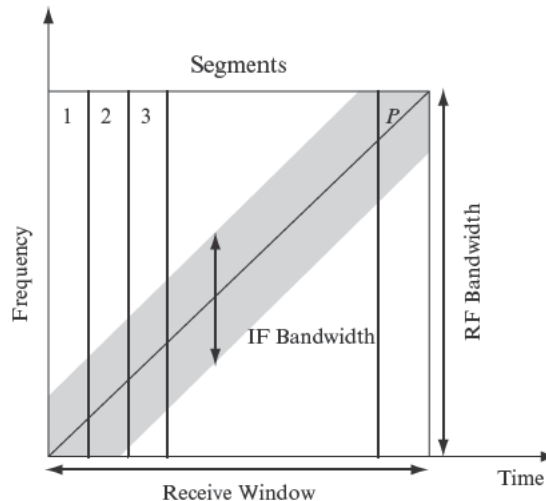
FIGURE 9-50 ■
Stretch processor
response to a white
noise jammer.



The first sinc term in this equation is due to the wideband dispersion of the jamming signal over the intermediate frequency bandwidth caused by the spatial propagation delay. The second periodic sinc term is decorrelation due to the nonstationarity from the stretch processing. As K increases, that periodic sinc term becomes small, thereby reducing the spatial correlation of the jammer and degrading the ability of the adaptive beamformer to cancel it. To avoid this degradation, the adaptive weights must be updated within the receive window so that they can evolve to keep up with the variation of the modulated jammer signal over time [40]. A simple way to implement this updating is to subdivide the receive window into segments and then recompute the adaptive weights within each segment using some or all of the segment data to train the weights, as shown in Figure 9-51. For the special case of poststretch cancellation, breaking up the receive window into segments is analogous to subbanding.

Updating the weights within the receive window also has some potential drawbacks. The number of data samples available for estimating the adaptive weights is limited by

FIGURE 9-51 ■
Updating adaptive
weights within the
receive window for
poststretch jammer
cancellation.



the length of the segment and may not provide adequate sample support. This creates a trade-off between cancellation performance loss due to estimation errors, which motivates making the segments larger, and performance loss due to the jammer nonstationarity, which motivates making the segments smaller. A second drawback is that changing the weights periodically in the receive window can impose small modulations on the target signal that create increased range sidelobes after pulse compression.

9.6 | SUMMARY

DBF architectures provide significant functionality enhancements for phased array radar. At the current state of digital technology, element-level DBF is often impractical for large arrays operating at high frequencies, so subarray level implementations are commonly used to reduce the digital receiver count. The choice of subarray architecture heavily impacts performance due to grating lobes and grating nulls that result from under sampling the array. Adaptive algorithms provide the ability to cancel unwanted jamming and can use low gain auxiliaries, subarrays, and high gain beams as spatial degrees of freedom; each of these work best against jamming located in specific regions of the antenna pattern. The adaptive filters can also be augmented for wideband operation by optimize over both the spatial and frequency domains.

9.7 | FURTHER READING

Introductory material on adaptive filtering is available in [1,8].

9.8 | REFERENCES

- [1] Johnson, D.H. and Dudgeon, D.E., *Array Signal Processing: Concepts and Techniques*, Prentice Hall, Englewood Cliffs, NJ, 1993.
- [2] Haykin, S. and Steinhardt, A., Ed., *Adaptive Radar Detection and Estimation*, Wiley, New York, 1992.
- [3] Hudson, J.E., *Adaptive Array Principles*, Peter Peregrinus Ltd., London, 1989.
- [4] Monzingo, R.A. and Miller, T.W., *Introduction to Adaptive Arrays*, Wiley, New York, 1980.
- [5] Compton, R.T., *Adaptive Antennas: Concepts and Performance*, Prentice Hall, Englewood Cliffs, NJ, 1988.
- [6] Widrow, B. and Stearns, S., *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1985.
- [7] Van Trees, H.L., *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*, Wiley, New York, 2002.
- [8] Sayed, A.H., *Fundamentals of Adaptive Filtering*, IEEE Press, John Wiley & Sons, 2003.
- [9] Mailloux, R.J., *Phased Array Antenna Handbook*, 2d ed., Artech House, Boston, MA, 2005.
- [10] Haykin, S., *Adaptive Filter Theory*, 4th ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.
- [11] Laakso, T.I., Valimäki, V., Karjalainen, M., and Laine, U.K., "Splitting the Unit Delay," *IEEE Signal Processing Magazine*, pp. 30–60, January 1996.

- [12] Rabideau, D.J., "Hybrid Mitigation of Distortion in Digital Arrays," in *IEEE International Radar Conference*, pp. 236–241, May 9–12, 2005.
- [13] Lauritzen, K.C., et al., "High-Dynamic-Range Receivers for Digital Beam Forming Radar Systems," In *Proceedings of the IEEE Radar Conference*, Boston, MA, pp. 55–60, April 17–20, 2007.
- [14] Van Veen, B., "An Analysis of Several Partially Adaptive Beamformer Designs," *IEEE Transactions on ASSP*, vol. 37, no. 2, pp. 192–203, February 1989.
- [15] Nickel, U., "Subarray Configurations for Interference Suppression with Phased Array Radar," in *Proceedings of the International Conference on Radar*, Paris, pp. 82–86, 1989.
- [16] Johannisson, B. and Steen, L., "Partial Digital Beamforming Using Antenna Subarray Division," in *Proceedings of the International Conference on Radar*, Paris, pp. 63–67, 1994.
- [17] Ferrier, J., Carrara, B., and Granger, P., "Antenna Subarray Architectures and Anti-jamming Constraints," in *Proceedings of the International Conference on Radar*, Paris, pp. 466–469, 1994.
- [18] Nickel, U., "Subarray Configurations for Digital Beamforming with Low Sidelobes and Adaptive Interference Suppression," in *Proceedings of the IEEE International Radar Conference*, pp. 714–719, 1995.
- [19] Combaud, M., "Adaptive Processing at the Subarray Level," *Aerospace Science and Technology*, no. 2, pp. 93–105, 1999.
- [20] Zatman, M., "Digitization Requirements for Digital Radar Arrays," in *Proceedings of the IEEE Radar Conference*, pp. 163–168, May 1–3, 2001.
- [21] Bailey, C.D. and Aalfs, D.D., "Design of Digital Beamforming Subarrays for a Multifunction Radar," in *Proceedings of the International Radar Conference*, Bordeaux, France, October 2009.
- [22] Lin, C.-T. and Ly, H., "Sidelobe Reduction through Subarray Overlapping for Wideband Arrays," *Proceedings of the IEEE Radar Conference*, May 1–3, 2001.
- [23] Brennan, L.E. and Reed, I.S., "Theory of Adaptive Radar," *IEEE Transactions on AES*, vol. 9, pp. 237–252, March 1973.
- [24] Frost, O.L., "An Algorithm for Linearly Constrained Adaptive Array Processing," *Proceedings of the IEEE*, vol. 60, pp. 124–133, August 1972.
- [25] Griffiths, L.J. and Jim, C.W., "An Alternative Approach to Linearly Constrained Adaptive Beamforming," *IEEE Transactions AP-30*, pp. 27–34, January 1982.
- [26] Er, M.H. and Cantoni, A., "Derivative Constraints for Broad-Band Element Space Antenna Array Processors," *IEEE Transactions on ASSP*, vol. 31, pp. 1378–1393, December 1983.
- [27] Buckley, K.M. and Griffiths, L.J., "An Adaptive Generalized Sidelobe Canceller with Derivative Constraints," *IEEE Transactions on AP*, vol. 34, pp. 311–319, March 1986.
- [28] Buckley, K.M., "Spatial/Spectral Filtering with Linearly Constrained Minimum Variance Beamformers," *IEEE Trans. ASSP*, vol. 35, pp. 249–266, March 1987.
- [29] Reed, I.S., Mallett, J.D., and Brennan, L.E., "Rapid Convergence Rate in Adaptive Arrays," *IEEE Transactions on AES*, vol. 10, pp. 853–863, November 1974.
- [30] Boroson, D.M., "Sample Size Considerations for Adaptive Arrays," *IEEE Transactions on AES*, vol. 16, pp. 446–451, July 1980.
- [31] Carlson, B.D., "Covariance Matrix Estimation Errors and Diagonal Loading in Adaptive Arrays," *IEEE Transactions on AES*, vol. 24, pp. 397–401, July 1988.

- [32] Golub, G.H. and Van Loan, C.F., *Matrix Computations*, 3d ed., Johns Hopkins University Press, 1996.
- [33] Anderson, E., et al., *LAPACK Users' Guide*, 3d ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999.
- [34] Rader, C.M., "VLSI Systolic Arrays for Adaptive Nulling," *IEEE Signal Processing Magazine*, July 1996.
- [35] Farina, A., *Antenna-Based Signal Processing Techniques for Radar Systems*, Artech House, Boston, MA, 1992.
- [36] Special Issue on Adaptive Antennas, *IEEE Transactions on AP*, vol. 24, September 1976.
- [37] Weiss, s., et al., "An Efficient Scheme for Broadband Adaptive Beamforming," in *Asilomar Conf. Sig. Sys. Comp.*, Monterey, CA, November 1999.
- [38] Vaidyanathan, P.P. "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial," *Proceedings of the IEEE*, vol. 78, no. 1, pp. 56–93, January 1990.
- [39] Fliege, N.J., *Multirate Digital Signal Processing: Multirate Systems—Filter Banks—Wavelets*, Wiley, New York, 1994.
- [40] Aalfs, D.D., Brown, G.C., Holder, E.J., and Beason, B., "Adaptive Beamforming for Wideband Radars," in *Proceedings of the International Radar Symposium*, Munich, Germany, pp. 983–992, September 15–27, 1998.

9.9 | PROBLEMS

1. What is the equation for the availability of an N -channel DBF system that can tolerate the failure of any two receivers? What is the general form of the equation for availability where P is the number of failed receivers that can be tolerated?
2. The (3:1) overlapped subarray layout for the 64-element linear array used to generate Figures 9-17 and 9-18 has six subarrays instead of eight. Why?
3. How could the example array from problem 2 be changed so that it will have eight 3:1 overlapped subarrays?
4. Derive the Wiener filter solution for the optimum adaptive weights by minimizing the mean squared error of the beamforming output.
5. Derive the maximum SINR solution for the optimum adaptive weights.
6. Derive the LCMV solution using the method of Lagrange multipliers.
7. According to the principle of orthogonality, the output, $y(n)$, of the optimal filter is orthogonal to the error (i.e., $E[y(n)e^*(n)] = 0$). Use this relationship to derive the optimal solution for the weights.
8. Write an expression for the covariance matrix of a plane wave in the presence of noise that is independent and identically distributed. The expression should be in terms of the steering vector for the plane wave signal, the signal power, and the noise power.
9. Using the matrix inversion lemma, write an expression for the inverse of the covariance matrix from problem 6.
(Matrix inversion lemma: a matrix of the form $A = B^{-1} + CD^{-1}C^H$ has an inverse given by $A^{-1} = B + BC(D + C^HBC)^{-1}C^HB$).

10. In Figure 9-25, why is the max SINR value for the two-element array equal to 3 dB?
11. Derive the equation for the instantaneous cross-correlation of a white noise jammer after stretch processing that was given in Section 9.5.4.
12. Verify the equation given in Section 9.5.4 for jammer cross-correlation averaged over K time samples.

Clutter Suppression Using Space-Time Adaptive Processing

William L. Melvin

Chapter Outline

10.1	Introduction	453
10.2	Space-Time Signal Representation	459
10.3	Space-Time Properties of Ground Clutter	472
10.4	Space-Time Processing	474
10.5	STAP Fundamentals	478
10.6	STAP Processing Architectures and Methods	483
10.7	Other Considerations	491
10.8	Further Reading	493
10.9	Summary	493
10.10	References	494
10.11	Problems	496

10.1 INTRODUCTION

Moving target indication (MTI) is a common radar mission involving the detection of airborne or surface moving targets. The signal-to-noise ratio (SNR)—a characterization of the noise-limited performance of the radar against a target with radar cross section σ_T at range r —is approximated as

$$SNR(\phi, \theta) = \left(\frac{P_t G_t(\phi, \theta)}{4\pi r^2} \right) \left(\frac{\sigma_T}{4\pi r^2} \right) \left(\frac{A_e G_{sp}}{N_{in} F_n L_{rf}} \right) \quad (10.1)$$

where P_t is peak transmit power, $G_t(\phi, \theta)$ is antenna gain at azimuth ϕ and elevation θ , A_e is the effective receive aperture area, G_{sp} represents processing gains, N_{in} is the input noise power, F_n is the receiver noise figure, and L_{rf} represents radio frequency (RF) system losses [1]. Assuming the noise is uncorrelated (white) and Gaussian, the probability of detection is a one-to-one, monotonic function of both SNR and probability of false alarm: by maximizing SNR , the processor maximizes probability of detection for a fixed probability of false alarm. In light of (10.1), the radar designer ensures detection of targets with diminishing radar cross section at farther range by increasing the power-aperture product, $P_t A_e$. System constraints and cost limit the deployable power-aperture product.

The aerospace radar system design must also accommodate the impact of ground clutter on moving target detection. Clutter returns can appear significantly stronger than the target response. Unlike white noise (uncorrelated receiver noise or sky noise), clutter exhibits a degree of correlation structure evident in the frequency domain. Strong clutter returns increase the amount of ambiguity in the target decision process, making it more difficult to decide whether a target is present in a given observation. Analogous to the white noise detection scenario, the probability of detection depends on both signal-to-clutter-plus-noise ratio and the specified false alarm rate. It is common to collectively refer to clutter returns and other correlated sources of noise as interference and evaluate the signal-to-interference-plus-noise ratio (SINR) as one of the key factors influencing detection performance. Since $SINR \leq SNR$, interference always degrades detection performance compared with the noise-limited case.

Signal diversity, in the form of spatial and temporal degrees of freedom (DoFs), greatly enhances radar detection in the presence of certain types of colored noise. Specifically, the appropriate application of space-time DoFs maximizes SINR when the target competes with ground clutter. Clutter exhibits correlation in both spatial and (pulse-to-pulse) temporal dimensions. Space-time adaptive processing (STAP) involves adaptively (or dynamically) adjusting a two-dimensional space-time filter response in an attempt at maximizing the filter's output SINR and, consequently, improving radar detection performance.

The objective of this chapter is to develop the basic theory of STAP as it relates to aerospace radar detection of moving targets in clutter-limited environments. STAP is a higher dimensional version of the adaptive sidelobe canceller developed in the late 1950s and early 1960s [2]. Brennan and Reed discuss the theory of STAP for airborne radar in a seminal 1973 paper [3]. In years since the Brennan and Reed paper, STAP has been vigorously researched [4–8]. The recent advancement of high-speed, high-performance, digital signal processors makes fielding STAP-based radar systems possible on manned and unmanned airborne platforms and satellites.

10.1.1 STAP Overview

STAP is an adaptive filtering technique using spatial and temporal DoFs. A multichannel array is used to measure the spatial characteristics of a propagating electromagnetic wave; the Fourier transform of space is angle. Radar system designers recognize two primary time measurements: fast time and slow time. Fast time corresponds to the interval during which the receiver ingests reflected, pulsed signals as they return from specific distances relative to the radar. Slow time is the interpulse time; its Fourier transform yields Doppler frequency. The clutter signal exhibits correlation in the spatial and slow-time dimensions and generally appears white in fast time.

In the case of aerospace radar clutter mitigation, the STAP attempts to drive a deep null at the angle-Doppler frequencies coinciding with the clutter signal's spectral support while maximizing the receive gain at the target's angle-Doppler frequency. Stationary clutter scatterers fall along a ridge in the angle-Doppler domain because of the well-known coupling of a stationary scatterer's angular position with Doppler frequency. The STAP uses a space-time weighting incorporating information on the dynamically varying properties of the interference environment to drive the appropriate clutter null location, depth, and width. Maximizing output SINR is the primary objective.

It is known that maximizing SINR at the output of the space-time processor equivalently maximizes probability of detection for a fixed false alarm rate [3]. Most STAP performance metrics thus focus on characterizing algorithm impacts on SINR. SINR loss is generally spatially and temporally varying and characterizes SINR degradation due to the presence of ground clutter.

Reduced-dimension STAP (RD-STAP) techniques are common and use frequency-domain transforms and selection operators to reduce the dimensionality of the adaptive processor. Usually a small loss in performance relative to the joint space-time filter results when using RD-STAP. However, its benefits include significantly reduced computational loading and more efficient use of training data, often leading to more effective adaptive filter implementation.

Covariance estimation is a key consideration in the practical implementation of STAP. Heterogeneous or nonstationary clutter violates the requirement that training data used to estimate the unknown covariance matrix be homogeneous with respect to the null-hypothesis condition of the cell under test. Degraded output SINR and an increase in the detector's false alarm rate are the typical consequences. Robust STAP methods must make provision for heterogeneous or nonstationary clutter.

10.1.2 Organization

The subsequent sections introduce the reader to common mathematical operations and notation used throughout the chapter. STAP discussion relies extensively on the use of linear algebra to describe signal and filter characteristics.

After this preliminary discussion, this chapter develops spatial, temporal (Doppler), and space-time signal representation. Additionally, the early sections of this chapter introduce the reader to spatial beamforming, Doppler processing, and two-dimensional matched filtering, which are critical components of STAP. The space-time signal formulation is then used to develop a ground clutter signal model and corresponding clutter covariance matrix.

The middle sections of this chapter provide a general framework for space-time processing and introduce key STAP performance metrics, including probability of detection, probability of false alarm, SINR loss factors, and improvement factor. SINR loss is perhaps the most commonly used STAP metric, and two variants are typical: clairvoyant SINR loss, which compares the SINR at the output of a space-time filter with known weights to the achievable output SNR; and adaptive SINR loss, which compares the adaptive filter output SINR to the optimal (known weights) filter SINR. After a discussion of metrics, several methods used to calculate the optimal weight vector are given, including the maximum SINR filter, the minimum variance (MV) beamformer, and the generalized sidelobe canceller (GSC).

Most STAP discussion centers on ways to approximate the unknown, space-time weight vector and implement the filter. Two fundamental approaches exist and are discussed in the last third of the chapter. The first, and most popular, approach is known as reduced-dimension STAP, which applies deterministic filters and selection operations to reduce the dimensionality of the STAP filter and therefore the computational burden and requirements on training sample support. The second approach is reduced-rank STAP, a weight calculation method based on eigendecomposition of the interference-plus-noise covariance matrix. In the most general sense it results in a projection of the data to suppress interference, followed by a matched filtering operation. Since reduced-rank STAP

is actually a weight calculation method, it is compatible with reduced-dimension STAP. Post-Doppler, reduced-dimension STAP, as described herein, is especially attractive for implementation.

The last sections of this chapter consider a practical, end-to-end detection architecture incorporating STAP. Angle-Doppler estimation is discussed in this context. Some important, real-world issues concerning covariance matrix estimation challenges are given in the final section: we briefly define and discuss the implications of heterogeneous (culturally varying) and nonstationary (sensor-induced, angle-Doppler varying) clutter.

10.1.3 Key Points

- STAP is an adaptive implementation of an optimal, space-time filter.
- The STAP weight vector has two key components. The first is an inverse of the unknown, interference covariance matrix containing information on clutter and jamming that degrade detection performance; the inverse mitigates their influence. The second is the space-time steering vector used to integrate energy from a specified angle of arrival and Doppler frequency.
- The STAP uses training data to estimate unknown, second-order interference statistics to calculate the unknown weight vector.
- Due to computational loading and limited training data, reduced-dimension STAP is a practical approach suited to real-world implementation.
- A number of practical matters require careful consideration when implementing STAP, including methods to mitigate heterogeneous and nonstationary clutter.

10.1.4 Notation and Operations

In general, a boldface, lowercase variable indicates a vector quantity; a boldface, uppercase variable indicates a matrix; and a variable with a caret is an estimate. Superscripts T or H applied to a vector or matrix denote the transpose or Hermitian (conjugate) transpose operations.

The notation $\mathbf{a} \sim CN(\boldsymbol{\mu}_a, \mathbf{R}_a)$ indicates that \mathbf{a} is complex normal (Gaussian) with mean $\boldsymbol{\mu}_a$ and covariance matrix \mathbf{R}_a . The notation $\mathbf{x} \in C^{m \times 1}$ indicates \mathbf{x} is an element of the set of complex vectors of dimension $m \times 1$, while $\mathbf{X} \in C^{m \times n}$ indicates a matrix belonging to the set of complex matrices of dimension $m \times n$. Also, $[\mathbf{x}]_m$ or x_m is the m -th element of \mathbf{x} , $[\mathbf{X}]_{m,n}$ or $X_{m,n}$ is the $m - n$ -th element of \mathbf{X} , \mathbf{x}_k or \mathbf{X}_k represent the k -th realization of the given vector or matrix, and $\mathbf{x}_k(n)$ is the n -th sample of the k -th realization (e.g., $\mathbf{x}_k(n)$ might represent the vector of voltages measured at M antenna elements for the k -th range and n -th pulse). Also, $\{\mathbf{x}_m\}_{m=1}^P$ is the collection of column vectors, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_P$.

The following linear algebraic operations are used: inner product, outer product, Kronecker (tensor) product, and Schur (Hadamard) product. The inner product describes the output of a finite impulse response (FIR) filter. Given the weight vector $\mathbf{w} \in C^{P \times 1}$ and data vector $\mathbf{x} \in C^{P \times 1}$, the inner product defines the peak output of a P -tap FIR filter

$$\mathbf{w}^H \mathbf{x} = \begin{bmatrix} w_1^* & w_2^* & \cdots & w_P^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_P \end{bmatrix} = \sum_{m=1}^P [\mathbf{w}]_m^* [\mathbf{x}]_m \quad (10.2)$$

where $[\mathbf{w}]_m^* = w_m^*$ is the conjugate of the m -th element of \mathbf{w} , and $[\mathbf{x}]_m = x_m$ is the m -th element of \mathbf{x} . Optimization problems typically involve the expected value of the outer product; the outer product of the data vector is the $P \times P$ matrix

$$\mathbf{xx}^H = \begin{bmatrix} x_1 x_1^* & x_1 x_2^* & \cdots & x_1 x_P^* \\ x_2 x_1^* & x_2 x_2^* & \cdots & x_2 x_P^* \\ \vdots & \vdots & \ddots & \vdots \\ x_P x_1^* & x_P x_2^* & \cdots & x_P x_P^* \end{bmatrix} \quad (10.3)$$

The Kronecker product conveniently describes a collection of multidimensional observations. For example, given vectors $\mathbf{s}_s \in C^{M \times 1}$ and $\mathbf{s}_t \in C^{N \times 1}$, the Kronecker product is

$$\mathbf{s}_t \otimes \mathbf{s}_s = [\mathbf{s}_t]_1 \mathbf{s}_s^T \quad [\mathbf{s}_t]_2 \mathbf{s}_s^T \quad \cdots \quad [\mathbf{s}_t]_N \mathbf{s}_s^T]^T \in C^{NM \times 1} \quad (10.4)$$

where $[\mathbf{s}_t]_m$ is the m -th element of \mathbf{s}_t . Lastly, the Schur product is tantamount to element-wise multiplication of two vectors of the same length. For example, given vector $\mathbf{s}_s \in C^{M \times 1}$ and the window function (taper) $\mathbf{a}_s \in C^{M \times 1}$, the Schur product is

$$\mathbf{a}_s \odot \mathbf{s}_s = [\mathbf{a}_s]_1 [\mathbf{s}_s]_1 \quad [\mathbf{a}_s]_2 [\mathbf{s}_s]_2 \quad \cdots \quad [\mathbf{a}_s]_M [\mathbf{s}_s]_M]^T \quad (10.5)$$

The notation $:$ is taken from the MATLAB programming language. It is best understood by example. For instance

$$\mathbf{a} = [n_1 : \Delta : n_2] \quad (10.6)$$

is a row vector with the first element equal to n_1 , subsequent elements incremented by Δ , and the last element equal to n_2 . If $\Delta = 1$, it is usually dropped from (10.6), viz. $[n_1 : 1 : n_2] = [n_1 : n_2]$. For example, $\mathbf{a} = [-3.25 : 1.25 : 1.75] = [-3.25 \ -2 \ -0.75 \ 0.5 \ 1.75]$.

10.1.5 Variable Names

Frequently used variables are as follows:

- ω_c = center frequency (rads)
- f_d = Doppler frequency (Hz)
- \tilde{f}_d = normalized Doppler frequency (unitless)
- ϕ, θ = azimuth and elevation (rads)
- ϕ_c = cone angle (rads)
- λ_o = center wavelength (m)
- f_{sp} = spatial frequency (unitless)
- M = number of channels
- N = number of pulses
- L = number of available range bins
- T = pulse repetition interval (s)
- c = velocity of propagation (speed of light, m/s)
- σ_n^2 = receiver noise power, single channel/pulse (watts)
- σ_s^2 = signal power, single channel/pulse (watts)
- \mathbf{v}_p = platform velocity vector (m/s; m/s; m/s)
- \mathbf{s}_s = spatial steering vector
- \mathbf{v}_s = hypothesized spatial steering vector

\mathbf{s}_t = temporal steering vector
 \mathbf{v}_t = hypothesized temporal steering vector
 \mathbf{s}_{s-t} = space-time steering vector
 \mathbf{v}_{s-t} = hypothesized space-time steering vector
 $\mathbf{x}_{k/s}(n)$ = spatial data snapshot, k -th range cell, n -th pulse
 \mathbf{x}_k = space-time data snapshot, k -th range cell
 \mathbf{c}_k = clutter space-time snapshot, k -th range cell
 \mathbf{j}_k = RFI space-time snapshot, k -th range cell
 \mathbf{n}_k = uncorrelated noise space-time snapshot, k -th range cell
 $\bar{\mathbf{s}}_k$ = target space-time snapshot, k -th range cell
 \mathbf{R}_k = null-hypothesis covariance matrix, k -th range cell
 $\mathbf{R}_{c/k}$ = clutter covariance matrix
 $\hat{\mathbf{R}}_k$ = null-hypothesis covariance estimate, k -th range cell
 \mathbf{R}_s = signal covariance matrix
 $\mathbf{q}_k(m)$ = m -th eigenvector for covariance matrix, k -th range cell
 $\lambda_k(m)$ = m -th eigenvalue for covariance matrix, k -th range cell
 r_c = clutter covariance matrix rank
 \mathbf{I}_p = P -by- P identity matrix
 \mathbf{w}_s = spatial weight vector
 \mathbf{w}_t = temporal weight vector
 \mathbf{w}_k = space-time weight vector, k -th range
 $\hat{\mathbf{w}}_k$ = adaptive space-time weight vector, k -th range
 P_D = probability of detection
 P_{FA} = probability of false alarm.

10.1.6 Acronyms

ADC	analog-to-digital converter
AMF	adaptive matched filter
CFAR	constant false alarm rate
CNR	clutter-to-noise ratio
CPI	coherent processing interval
DOA	direction of arrival
DPCA	displaced phase center antenna processing
DoF	degree of freedom
EFA	extended factored algorithm
FFT	fast Fourier transform
FLA	forward-looking array
GSC	generalized sidelobe canceller
IF	improvement factor
IID	independent and identically distributed
INR	interference-to-noise ratio
JDL	joint domain localized
KLT	Karhunen-Loève transform
MCARM	Multi-Channel Airborne Radar Measurements
MDV	minimum detectable velocity
MMSE	minimum mean square error
MTI	moving target indication

MV	minimum variance
MVDR	minimum variance distortionless response
PCI	principal components inverse
PDI	post-detection integration
PSD	power spectral density
PRF	pulse repetition frequency
PRI	pulse repetition interval
RD-STAP	reduced-dimension STAP
RF	radio frequency
RR-STAP	reduced-rank STAP
Rx	receiver
SNR	signal-to-noise ratio
SINR	signal-to-interference-plus-noise ratio
SLA	side-looking array
STAP	space-time adaptive processing
ULA	uniform linear array
UDSF	usable Doppler space fraction

10.2 | SPACE-TIME SIGNAL REPRESENTATION

In this section we describe spatial, temporal, and space-time signal representation. Since MTI radar is our primary concern, we consider temporal signal properties in the context of slow-time (pulse-to-pulse) sampling. We begin by describing spatial sampling using a multichannel array, along with the corresponding spatial filtering operation known as beamforming. Thereafter, we discuss temporal sampling and Doppler processing as mathematical extensions of the spatial case. We then extend the separate spatial and temporal developments to describe simultaneous space-time signals.

10.2.1 Spatial Sampling and Beamforming

Using an array of spatial sensors, the radar determines the direction of a propagating electromagnetic wave by effectively measuring time difference of arrival as a phase variation across the aperture [2,9]. Specifically, the array digitizes the voltage outputs of M channels, thereby gathering spatial information embodied in the relative phase of each of the measured signals. A collection of antenna elements, known as a subarray, feeds each receiver channel. In our discussion we assume the following conditions: the far-field approximation applies, indicating a planar, propagating wave; the signals are narrowband, suggesting the signal bandwidth is a very small fraction of the signal carrier frequency; M equal size subarrays comprise the overall antenna array; and the array configuration exhibits uniform spacing with horizontal, linear orientation. Figure 10-1 depicts the corresponding scenario of a propagating plane wave impinging on a uniform linear array (ULA).

The signal direction of arrival (DOA), frequency, and physical location of a given subarray influence the varying phase shift the propagating wave induces among each of the channels. Letting $\tau_{s,m}$ represent the delay between the time it takes the wavefront to arrive at a suitably designated reference point and the m -th subarray and defining ω_c as the center frequency of the propagating signal in radians, the corresponding phase shift is

$$\gamma_{s,m} = \tau_{s,m}\omega_c \quad (10.7)$$

the physical center of the m -th subarray as

$$\mathbf{d}_{s,m} = d_{x,m}\hat{\mathbf{x}} + d_{y,m}\hat{\mathbf{y}} + d_{z,m}\hat{\mathbf{z}}, \quad (10.9)$$

where $\{d_{x,m}, d_{y,m}, d_{z,m}\}$ describes the subarray location in Cartesian coordinates and $m = 1, 2, \dots, M$. Also, in this coordinate system the origin serves as the reference point. Given (10.8) and (10.9), the distance the waveform must travel to reach the m -th sensor, with respect to the reference point, is

$$d_{r,m} = \mathbf{k}(\phi, \theta) \cdot \mathbf{d}_{s,m} = \begin{bmatrix} \cos \theta \sin \phi & \cos \theta \cos \phi & \sin \theta \end{bmatrix} \begin{bmatrix} d_{x,m} \\ d_{y,m} \\ d_{z,m} \end{bmatrix} \quad (10.10)$$

For the special case of a ULA oriented along the x-axis

$$\mathbf{d}_{s,m} = [\tilde{\mathbf{n}}]_m d \hat{\mathbf{x}} \quad (10.11)$$

defines the subarray location, where $[\tilde{\mathbf{n}}]_m$ is the m -th element of the vector $\tilde{\mathbf{n}}$

$$\tilde{\mathbf{n}} = \left[\frac{1-M}{2} : \frac{M-1}{2} \right] \quad (10.12)$$

and d is the spacing between subarrays.

The angle ϕ_c identified in Figure 10-1 and Figure 10-2 is known as cone angle. A ULA does not uniquely measure azimuth and elevation; rather, it senses cone angle. In this case, the cone angle defines a conical ambiguity surface centered on the x-axis with the apex at the origin. Signals emanating from different azimuths and elevations, but with the same cone angle, are indistinguishable. Using (10.11) in (10.10) gives

$$d_{r,m} = [\tilde{\mathbf{n}}]_m d \cos \theta \sin \phi \quad (10.13)$$

We recognize $\cos \theta \sin \phi$ as the direction cosine defining the angle from the x-axis to the unit vector $\mathbf{k}(\phi, \theta)$, implying $\cos \phi_c = \cos \theta \sin \phi$. Using the cone angle measurement

$$d_{r,m} = [\tilde{\mathbf{n}}]_m d \cos \phi_c \quad (10.14)$$

From (10.14) we find a broadside signal ($\phi_c = \pi/2$) indicates $d_{r,m} = 0$ for all m , so the relative phase is constant among all channels. On the other hand, $\phi_c = 0$ for an end-fire signal, meaning $d_{r,m} = [\tilde{\mathbf{n}}]_m d$.

Based on the preceding discussion, the phase at the m -th element for the ULA oriented along the x-axis is

$$\gamma_{s,m} = \tau_{s,m} \omega_c = \frac{\mathbf{k}(\phi, \theta) \cdot \mathbf{d}_{s,m}}{c} \omega_c = \frac{2\pi}{\lambda_o} [\tilde{\mathbf{n}}]_m d \cos \phi_c \quad (10.15)$$

where $\omega_c/c = 2\pi/\lambda_o$, and λ_o is the center wavelength. The signal at the m -th subarray for a given sampling instant t_s is

$$r_{s/m}(t_s) = a(t_s) \exp(j2\pi f_{sp} [\tilde{\mathbf{n}}]_m) \exp(j\omega_c t_s) \quad (10.16)$$

where $a(t_s)$ is the target complex envelope, and f_{sp} is known as spatial frequency. The first complex exponential term describes the channel-dependent phase due to signal DOA, whereas the latter exponential term characterizes the carrier signal. As the name suggests,

spatial frequency characterizes the spatially varying phase across the antenna aperture, just as temporal frequency characterizes temporally varying phase; the expression for spatial frequency is given by

$$f_{sp} = \frac{d}{\lambda_o} \cos \phi_c \quad (10.17)$$

After demodulation, the spatial snapshot (spatial data vector) follows from (10.16) as the length- M vector of complex, baseband voltages received by the various spatial channels:

$$\mathbf{x}_s(t_s) = a(t_s) \left[e^{j2\pi f_{sp}(\frac{1-M}{2})} \quad e^{j2\pi f_{sp}(\frac{3-M}{2})} \quad \dots \quad e^{j2\pi f_{sp}(\frac{M-5}{2})} \quad e^{j2\pi f_{sp}(\frac{M-3}{2})} \quad e^{j2\pi f_{sp}(\frac{M-1}{2})} \right]^T \quad (10.18)$$

The *spatial steering vector* characterizes the relative signal phase among channels and follows from (10.18) as

$$\mathbf{s}_s(f_{sp}) = \left[e^{j2\pi f_{sp}(\frac{1-M}{2})} \quad e^{j2\pi f_{sp}(\frac{3-M}{2})} \quad \dots \quad e^{j2\pi f_{sp}(\frac{M-5}{2})} \quad e^{j2\pi f_{sp}(\frac{M-3}{2})} \quad e^{j2\pi f_{sp}(\frac{M-1}{2})} \right]^T \quad (10.19)$$

Assigning the index k to the discrete sampling time t_s , we may then compactly express the spatial snapshot of a signal emanating from cone angle ϕ_c with complex amplitude and phase given by $a_k = a(t_s)$ as $\mathbf{x}_{k/s} = a_k \mathbf{s}_s(f_{sp})$.

In the beginning of this section we set forth the narrowband approximation as an underlying assumption. Consequently, we replace the precise wavelength λ by the carrier's center value λ_o . Examining (10.15), we observe that spatial phase is a function of frequency; as the signal bandwidth increases, phase variation with frequency becomes noticeable. Hence, a single signal arriving from angle ϕ_c begins to appear like multiple signals with DOAs centered on ϕ_c . This effect is known as dispersion. Additionally, the narrowband assumption is critical in assuming the signal's complex envelope is stationary over the aperture during the sample time; if the complex envelope changes over the array during the sample time, which is likely to happen only for very specialized applications, compensation techniques are necessary.

A beamformer, or spatial matched filter, is a band-pass filter exhibiting maximum gain at a particular DOA or look angle. The beamformer implementation involves weighting and combining the individual spatial samples. By definition, the spatial matched filter maximizes output SNR. Define $\mathbf{w}_s \in C^{M \times 1}$ as the spatial weight vector. The beamformer output is the scalar

$$y_k = \mathbf{w}_s^H \mathbf{x}_{k/s} = \sum_{m=1}^M [\mathbf{w}_s]_m^* [\mathbf{x}_{k/s}]_m \quad (10.20)$$

In the noise-only case, the power at the beamformer output is

$$P_n = E[\mathbf{w}_s^H \mathbf{n}_{k/s} \mathbf{n}_{k/s}^H \mathbf{w}_s] = \mathbf{w}_s^H \mathbf{R}_n \mathbf{w}_s \quad (10.21)$$

where $\mathbf{n}_{k/s} \sim CN(\mathbf{0}, \mathbf{R}_n)$ is the noise-only spatial snapshot, $\mathbf{R}_n = E[\mathbf{n}_{k/s} \mathbf{n}_{k/s}^H] = \sigma_n^2 \mathbf{I}_M$ is the noise covariance matrix, and σ_n^2 is the single-channel noise power. The output power

under the signal-only condition takes the form

$$P_s = E \left[\mathbf{w}_s^H \mathbf{x}_{k/s} \mathbf{x}_{k/s}^H \mathbf{w}_s \right] = \sigma_s^2 |\mathbf{w}_s^H \mathbf{s}_s(f_{sp})|^2 \quad (10.22)$$

where $\sigma_s^2 = E[|a_k|^2]$ is the single-channel signal power. Consequently, the output SNR is

$$SNR = \frac{P_s}{P_n} = \frac{\sigma_s^2 |\mathbf{w}_s^H \mathbf{s}_s(f_{sp})|^2}{\sigma_n^2 \mathbf{w}_s^H \mathbf{w}_s} \quad (10.23)$$

σ_s^2/σ_n^2 is the single-channel SNR. Since

$$\frac{|\mathbf{w}_s^H \mathbf{s}_s(f_{sp})|^2}{\mathbf{w}_s^H \mathbf{w}_s} \leq \frac{(\mathbf{w}_s^H \mathbf{w}_s) (\mathbf{s}_s^H(f_{sp}) \mathbf{s}_s(f_{sp}))}{\mathbf{w}_s^H \mathbf{w}_s} \quad (10.24)$$

we find that by setting $\mathbf{w}_s = \mathbf{s}_s(f_{sp})$, the left side of the inequality equals the right, and hence SNR is maximal. The expression in (10.24) is known as the Schwarz inequality. With this weight vector selection, (10.23) reduces to $SNR = (\sigma_s^2/\sigma_n^2) M$; in this instance we recognize M as the spatial integration gain.

While $\mathbf{w}_s = \mathbf{s}_s(f_{sp})$ defines the matched filter, oftentimes the processor applies a spatial taper to better control filter sidelobes at the expense of a widened main lobe and reduced SNR. If \mathbf{b}_s is the aperture apodization—a real valued function, such as a Taylor or Hanning weighting—then the weighted beamformer weight vector is given by $\mathbf{w}_s = \mathbf{b}_s \odot \mathbf{s}_s(f_{sp})$.

We also assumed an error-free array manifold in our discussion. Practical hardware mismatches among individual receive channels lead to random amplitude and phase errors. Consider the length- M random error vector $\boldsymbol{\varepsilon}$ whose m -th entry, $\varepsilon_m = [\boldsymbol{\varepsilon}]_m$, characterizes the amplitude and phase mismatch from the ideal value. In the presence of errors, the signal vector takes the form

$$\tilde{\mathbf{s}}_s = \boldsymbol{\varepsilon} \odot \mathbf{s}_s = \left[\varepsilon_1 e^{j2\pi f_{sp}(\frac{1-M}{2})} \quad \varepsilon_2 e^{j2\pi f_{sp}(\frac{3-M}{2})} \quad \dots \quad \varepsilon_{M-1} e^{j2\pi f_{sp}(\frac{M-3}{2})} \quad \varepsilon_M e^{j2\pi f_{sp}(\frac{M-1}{2})} \right]^T. \quad (10.25)$$

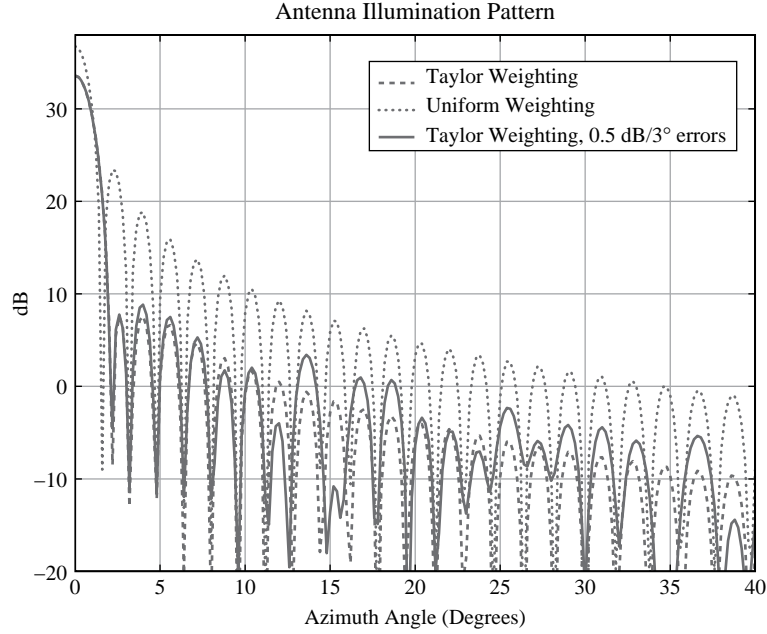
Array errors increase the sidelobe response of the spatial filter and degrade SNR due to the filter mismatch.

As an example, Figure 10-3 shows the beampattern for a 72-element ULA for uniform (unweighted) and Taylor weighted cases (25 dB Taylor, $\bar{n} = 4$). The Taylor pattern exhibits better sidelobe performance, a widened main beam, and decreased gain in comparison with the uniformly weighted array with attendant 13 dB sidelobe response. Additionally, we include the impact of element-level errors on the beampattern for 0.5 dB root mean square (RMS) amplitude variation and 3° RMS phase error. In this case, we find the errors do not affect pointing direction or main lobe width but do yield spurious increases in the antenna pattern sidelobes.

The spatial resolution of the beamformer is commonly given by either the null-to-null or 3 dB beamwidths. Null-to-null beamwidth is given as

$$B_{null} = \frac{\lambda_o}{L_{array}} \times \frac{180}{\pi} \text{ degrees} \quad (10.26)$$

FIGURE 10-3 ■
Spatial beampattern
for uniform and
Taylor weighted
ULAs, including
influence of array
errors.



where L_{array} is the array length in meters (substitute the array width into (10.26) for elevation beamwidth). The 3 dB beamwidth is

$$B_{3-dB} = \frac{a_{sw}\lambda_o}{L_{array}} \text{ degrees} \quad (10.27)$$

a_{sw} is a beam-broadening factor; typical values are in the range of $58 \leq a_{sw} \leq 70$ [1].

10.2.2 Temporal Sampling and Doppler Processing

MTI radar systems transmit and receive a sequence of N pulses as a means of temporally sampling the signal environment in the slow-time or Doppler dimension. The analog-to-digital converters (ADCs) sample in the fast time, or range, dimension. The radar signal processor coherently combines the N pulses in a manner very similar to spatial beamforming; this operation is commonly called Doppler processing or Doppler beamforming. Doppler beamforming separates the target signal from stationary ground clutter returns at the same angle and also enables estimation of target radial velocity.

Consider a length- N periodic pulse train, with a pulse repetition interval (PRI) of T . Further assume the presence of a point scatterer, initially at range r_o from the antenna phase center reference point. Due to the relative motion between the scatterer and radar platform, an interpulse change in range of Δr is present. The coherent dwell is kept sufficiently short to maintain a constant radial velocity between radar and scatterer, so the interpulse change in range is constant-valued. With this in mind, the corresponding fast-time delay resulting from the motion between platform and scatterer is

$$\tau_n = \frac{2r_o + n2\Delta r}{c} \quad (10.28)$$

where n is the pulse index, and the factor of two accounts for two-way range; as in the spatial case, τ_n is simply distance divided by the velocity of propagation.

As a result of the time-varying change in range, a pulse-to-pulse phase rotation occurs leading to a Doppler frequency offset. The phase for the n -th pulse is

$$\phi_n = \tau_n \omega_c = 4\pi \left[\frac{r_o + n \Delta r}{\lambda_o} \right] \quad (10.29)$$

This expression neglects bandwidth-related dispersion. Since radian frequency is the time derivative of phase, the corresponding Doppler frequency is

$$f_d = \frac{1}{2\pi} \frac{d\phi_n}{dt} = \frac{2\Delta r}{\lambda_o \Delta t} = \frac{2v_r}{\lambda_o}. \quad (10.30)$$

v_r denotes the radial velocity component, sometimes also called line-of-sight velocity or range rate. Equation (10.30) indicates that motion along a radial line between radar and target leads to the Doppler frequency offset.

The resulting temporal snapshot (temporal signal vector) corresponding to a point scatterer with normalized Doppler $\tilde{f}_d = f_d T$ is $\mathbf{x}_t = a_t \mathbf{s}_t(\tilde{f}_d)$, where a_t is the signal complex envelope (taken as perfectly correlated over the coherent dwell) and

$$\mathbf{s}_t(\tilde{f}_d) = [1 \quad e^{j2\pi\tilde{f}_d} \quad e^{j2\pi(2\tilde{f}_d)} \quad \dots \quad e^{j2\pi(N-1)\tilde{f}_d}]^T \quad (10.31)$$

$\mathbf{s}_t(\tilde{f}_d)$ is known as the *temporal steering vector*. Comparing (10.31) and (10.19), we find a mathematical similarity between spatial and temporal responses after writing (10.19) as

$$\mathbf{s}_s(f_{sp}) = \bar{a} [1 \quad e^{j2\pi f_{sp}} \quad e^{j2\pi(2f_{sp})} \quad \dots \quad e^{j2\pi(M-2)f_{sp}} \quad e^{j2\pi(M-1)f_{sp}}]^T \quad (10.32)$$

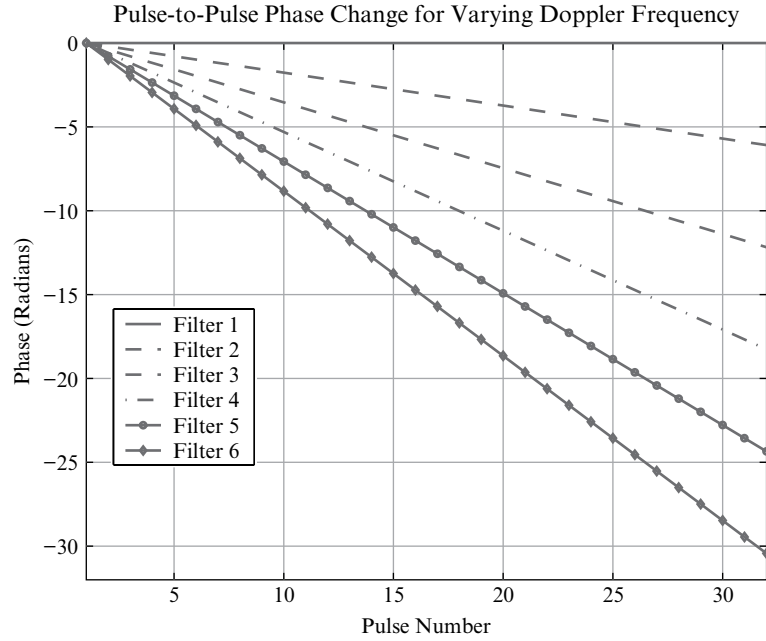
where \bar{a} is a complex scalar. For this reason, the collection of N receive pulses is sometimes called the temporal aperture.

Based on this similarity between spatial and temporal responses, we find by analogy to our discussion of the spatial matched filter in Section 10.2.1 that the temporal matched filter requires the weight vector $\mathbf{w}_t = \mathbf{s}_t(\tilde{f}_d)$. It is common to apply a real-valued weighting function to tailor the sidelobe response at the expense of broadened main lobe and loss of integration gain: $\mathbf{w}_t = \mathbf{b}_t \odot \mathbf{s}_t(\tilde{f}_d)$, where \mathbf{b}_t is a real-valued function, such as a Hanning or Chebyshev weighting.

Spatial or temporal matched filtering operations are coherent. The matched filter adjusts the phase of each sample and then sums the individual voltages. Referring to the collection of N pulses as the coherent processing interval (CPI)—common in MTI radar parlance—emphasizes this notion.

The temporal or spatial steering vectors in (10.31)–(10.32) assume a Vandermonde structure: each consecutive element of the vector is a multiple of the preceding term, increased by a single integer. This structure characterizes a linear phase ramp across the sampled aperture; the slope of the phase ramp corresponds to the particular frequency. Employing a fast Fourier transform (FFT) [9] enables the processor to efficiently test for the various phase ramp slopes, thereby accomplishing matched filtering. A weighting function is directly applied to the data in this instance. Straddle loss—the loss in SNR due to a mismatched filter response—occurs in practice since the precise target spatial or Doppler frequency is generally unknown. The difference between the precise linear phase

FIGURE 10-4 ■
Pulse-to-pulse
phase change over
32-pulse temporal
aperture for several
Doppler filters.



ramp and estimated slope, embodied in the FFT, leads to filter mismatch and consequent performance loss. Using our prior notation, it is common to replace $\mathbf{s}_t(\tilde{f}_d)$ and $\mathbf{s}_s(f_{sp})$ by the hypothesized responses $\mathbf{v}_t(\tilde{f}_d)$ and $\mathbf{v}_s(f_{sp})$. The processor will test the spatial and Doppler frequencies of interest for potential targets by stepping across various steering vectors $\mathbf{v}_t(\tilde{f}_d)$ and $\mathbf{v}_s(f_{sp})$. When employing uniform sampling (i.e., constant PRI or a ULA), the FFT efficiently accomplishes this task. A zero-padded FFT reduces straddle loss by decreasing the frequency bin spacing.

Figure 10-4 shows the pulse-to-pulse phase change over the temporal aperture for a 32-pulse CPI and several different normalized Doppler frequencies. The slope of each line is proportional to Doppler frequency, as (10.30) indicates. The response of each Doppler filter given by \mathbf{w}_t is reminiscent of the antenna beampatterns given in Figure 10-3. The dwell time, NT , determines the Doppler resolution (or Doppler beamwidth); the Doppler resolution is

$$\Delta_{f_d} = \frac{1}{\text{Dwell}} = \frac{1}{NT} = \frac{PRF}{N} \quad (10.33)$$

where the pulse repetition frequency (PRF) is the inverse of the PRI, $PRF = 1/T$.

10.2.3 Space-Time Signals

The preceding discussion of spatial and temporal sampling and signal structure is integral to a description of space-time signals. In this section we describe space-time signals and their characteristics in terms of power spectrum, covariance matrix, and eigenspectrum.

Suppose the array of sensors receives a signal with a particular Doppler frequency and direction of arrival. This signal could, for instance, result from a scattered transmission from a moving target. The space-time signal vector corresponding to the return from a

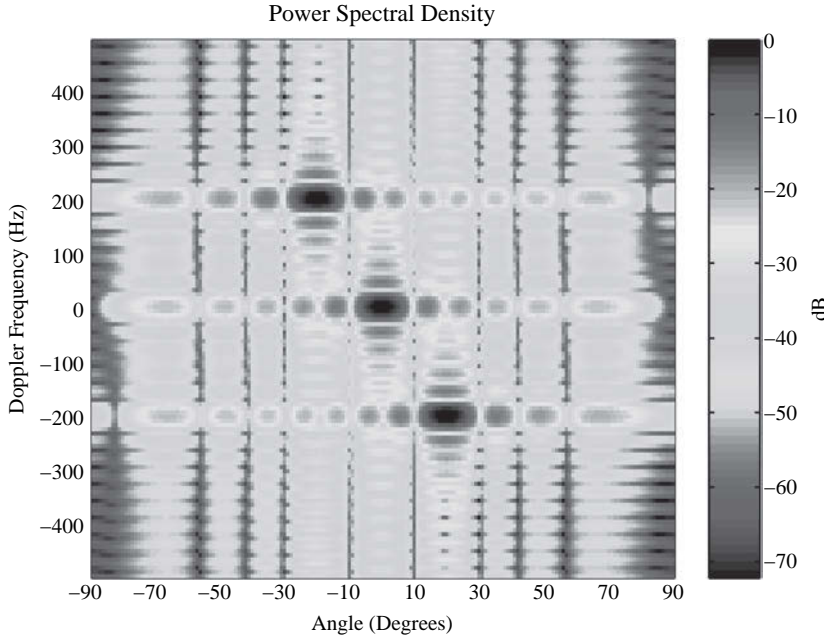


FIGURE 10-5 ■
Power spectral density of three space-time signals.

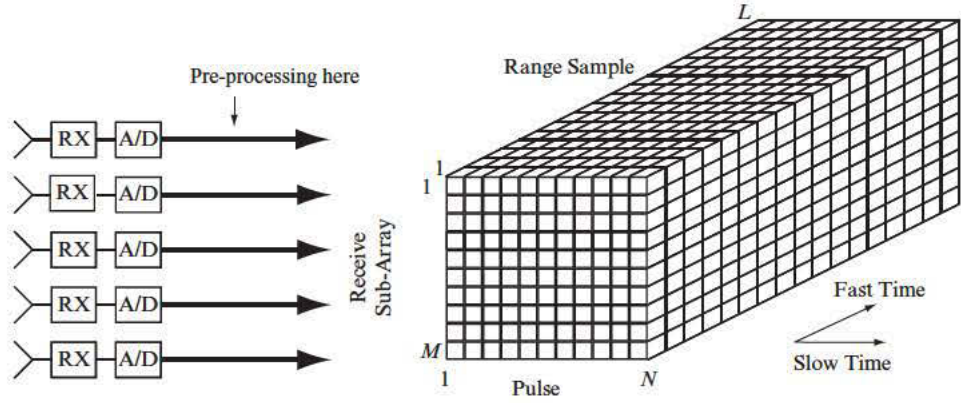
point scatterer with spatial frequency f_{sp} and normalized Doppler frequency \tilde{f}_d is $\mathbf{x}_{s-t} = a_{s-t} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$. Analogous to the one-dimensional spatial or temporal cases, a_{s-t} is a complex voltage proportional to the square root of the target radar cross section, and $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ is the $NM \times 1$ space-time steering vector. We conveniently describe the space-time steering vector as the Kronecker product of the temporal and spatial steering vectors:

$$\begin{aligned} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) &= \mathbf{s}_t(\tilde{f}_d) \otimes \mathbf{s}_s(f_{sp}) \\ &= \begin{bmatrix} 1 \cdot \mathbf{s}_s^T(f_{sp}) & e^{j2\pi \tilde{f}_d} \cdot \mathbf{s}_s^T(f_{sp}) & e^{j2\pi(2\tilde{f}_d)} \cdot \mathbf{s}_s^T(f_{sp}) \cdots e^{j2\pi(N-1)\tilde{f}_d} \cdot \mathbf{s}_s^T(f_{sp}) \end{bmatrix}^T. \end{aligned} \quad (10.34)$$

Figure 10-5 shows the two-dimensional power spectral density (PSD) of three unity amplitude space-time signals: the first signal corresponds to -200 Hz Doppler frequency and 20° DOA, the second to a Doppler of 200 Hz and DOA of -20° , and the third originating from array broadside with 0 Hz Doppler frequency. We use a 12-channel array ($M = 12$) receiving $N = 32$ pulses for this simulation. Observe that each space-time signal appears as a two-dimensional sinc function.

The multichannel, pulse-Doppler radar collects M spatial samples, N temporal samples, and L unambiguous range samples. The data cube of Figure 10-6, corresponding to a single CPI, organizes these samples. Preprocessing steps convert the RF signals at the multiple receiver channels to complex baseband space-time and range samples. The ADC clock rate is at least as high as the waveform bandwidth for complex sampling; the processor match filters the data in the range dimension and performs digital decimation (if necessary) to yield one complex, demodulated sample per range bin. The range dimension is known as fast time and the pulse dimension is called slow time since the sample rate is at the PRI and is much less than the ADC clock rate. The k -th space and slow-time slice

FIGURE 10-6 ■
Radar datacube.
(after [8], © 2003
IEEE)



of the data cube is

$$\mathbf{X}_k = \begin{pmatrix} [\mathbf{X}_k]_{1,1} & [\mathbf{X}_k]_{1,2} & \dots & [\mathbf{X}_k]_{1,N} \\ [\mathbf{X}_k]_{2,1} & [\mathbf{X}_k]_{2,2} & \dots & [\mathbf{X}_k]_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ [\mathbf{X}_k]_{M,1} & [\mathbf{X}_k]_{M,2} & \dots & [\mathbf{X}_k]_{M,N} \end{pmatrix} \quad (10.35)$$

The index k references the range cell of interest and $[\mathbf{X}_k]_{m,n}$ is the voltage corresponding to the m -th channel and n -th pulse. Vectorizing (10.35) by stacking each succeeding column one beneath the other yields the *space-time snapshot* for the k -th range, that is,

$$\mathbf{x}_k = [\mathbf{x}_{k/s}^T(1) \quad \mathbf{x}_{k/s}^T(2) \quad \dots \quad \mathbf{x}_{k/s}^T(N)]^T \quad (10.36)$$

where $\mathbf{x}_k \in \mathbb{C}^{MN \times 1}$ and $\mathbf{x}_{k/s}(n) \in \mathbb{C}^{M \times 1}$ is the spatial snapshot for the n -th pulse and k -th range. The operation leading to (10.36) is equivalent to the MATLAB programming language colon command: $\mathbf{x}_k = \mathbf{X}_k(:)$. Vectorizing (10.35) is convenient since we now can process the multidimensional signal in a manner analogous to the one-dimensional signal vector case; take advantage of linear algebraic notation and operations, such as the complex inner and outer products; and organize data for computational expediency on highly vectorized digital signal processing hardware.

The processor beamforms the data matrix of (10.35) to a single scalar representing a particular DOA and Doppler frequency by performing temporal matched filtering, or Doppler beamforming, across the N columns for a given spatial channel, thereby reducing the $M \times N$ matrix to an $M \times 1$ column vector and then spatially beamforming the remaining length- M data vector. For uniform space-time sampling, application of the two-dimensional FFT to (10.35) beamforms the data into multiple DOAs and Dopplers; the result is an $\tilde{M} \times \tilde{N}$ angle-Doppler matrix, where \tilde{M} and \tilde{N} represent the number of spatial and temporal points, respectively, in the 2-D FFT.

Analogous to the spatial and temporal cases, the space-time matched filter for snapshot $\mathbf{x}_k = a_{s-t} s_{s-t}(f_{sp}, \tilde{f}_d)$ is $\mathbf{w}_{k/s-t} = s_{s-t}(f_{sp}, \tilde{f}_d)$. The space-time matched filter maximizes the output SNR. Application of the 2-D FFT as a space-time beamformer approximates the matched filter over the measurable angle and Doppler space. As a means of tailoring the space-time sidelobes, the processor applies a two-dimensional weighting function to the k -th range sample via element-wise multiplication of (10.35) with $\mathbf{B}_{s-t} = \mathbf{b}_s \mathbf{b}_t^T$ (where \mathbf{b}_s and \mathbf{b}_t are the spatial and temporal weighting functions defined in our prior discussion).

The space-time covariance matrix describes all possible covariances among the elements of \mathbf{x}_k . Formally, its definition is

$$\mathbf{R}_k = E[\mathbf{x}_k \mathbf{x}_k^H] - \boldsymbol{\mu}_k \boldsymbol{\mu}_k^H; \boldsymbol{\mu}_k = E[\mathbf{x}_k] \quad (10.37)$$

$E[\cdot]$ is the expectation operator and $\boldsymbol{\mu}_k$ is the mean vector characterizing the mean value of each element of \mathbf{x}_k . We define the outer product operation in (10.3). Consider the case where $\boldsymbol{\mu}_k = \mathbf{0}$ (a typical, practical assumption). Then (10.37) takes the form

$$\mathbf{R}_k = \begin{bmatrix} E[|\mathbf{x}_k|_1|^2] & E[\mathbf{x}_k|_1 \mathbf{x}_k|_2^*] & \cdots & E[\mathbf{x}_k|_1 \mathbf{x}_k|_{MN}^*] \\ E[\mathbf{x}_k|_2 \mathbf{x}_k|_1^*] & E[|\mathbf{x}_k|_2|^2] & \cdots & E[\mathbf{x}_k|_2 \mathbf{x}_k|_{MN}^*] \\ \vdots & \cdots & \ddots & \vdots \\ E[\mathbf{x}_k|_{MN} \mathbf{x}_k|_1^*] & E[\mathbf{x}_k|_{MN} \mathbf{x}_k|_2^*] & \cdots & E[|\mathbf{x}_k|_{MN}|^2] \end{bmatrix} \quad (10.38)$$

The diagonal entries correspond to the variance of each element, $E[|\mathbf{x}_k|_m|^2] = E[\mathbf{x}_k|_m \mathbf{x}_k|_m^*]$, and the off-diagonal entries are the autocorrelations of varying lag [2,9].

Directly interpreting the significance of the covariance matrix—other than the diagonal entries, which serve as a measure of power—is difficult. The power spectral density is a more useful view of the properties of the signal environment. An important result relating the covariance matrix and PSD is

$$\mathbf{R}_k \stackrel{\mathfrak{F}_2}{\longleftrightarrow} \mathbf{S}(f_{sp}, \tilde{f}_d) \quad (10.39)$$

where \mathfrak{F}_p is the p -dimensional Fourier transform ($p = 2$ for the space-time case), and $\mathbf{S}(f_{sp}, \tilde{f}_d)$ is the PSD. The result in (10.39) follows from the Wiener-Khinchine theorem [9-11].

The periodogram is a biased estimate of the PSD [11]. Given the data cube of Figure 10-6, we compute the periodogram by averaging the 2-D Fourier Transform of individual snapshots. The estimate for $\mathbf{S}(f_{sp}, \tilde{f}_d)$ is thus given as

$$\hat{\mathbf{S}}(f_{sp}, \tilde{f}_d) = \frac{1}{K} \sum_{m=1}^K |\mathfrak{F}_2\{\mathbf{X}_m\}|^2. \quad (10.40)$$

Due to the presence of receiver noise, the covariance matrix is positive definite, viz.

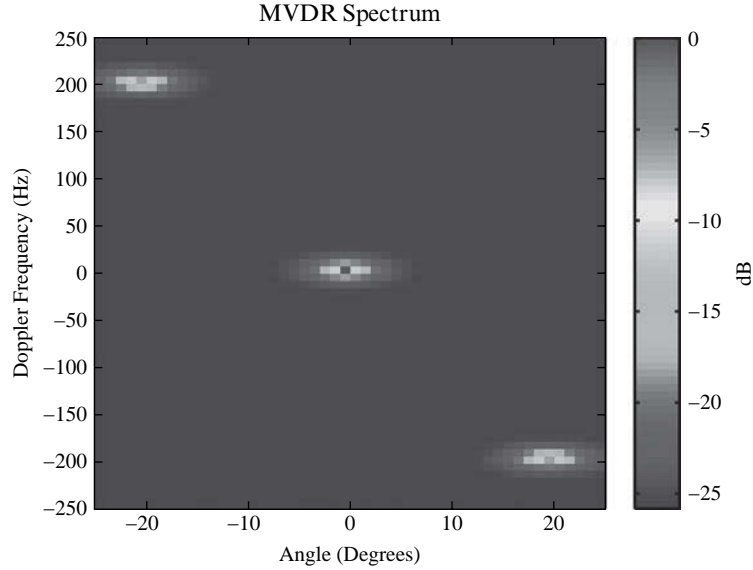
$$\mathbf{d}^H \mathbf{R}_k \mathbf{d} > 0 \quad (10.41)$$

for arbitrary $\mathbf{d} \neq \mathbf{0}$. We then can express the covariance matrix as [9]

$$\mathbf{R}_k = \sum_{m=1}^{NM} \lambda_k(m) \mathbf{q}_k(m) \mathbf{q}_k^H(m); \quad \lambda_k(1) \geq \lambda_k(2) \geq \cdots \lambda_k(NM) > 0 \quad (10.42)$$

$\lambda_k(m)$ is the m -th eigenvalue corresponding to eigenvector $\mathbf{q}_k(m)$. The collection of all eigenvectors forms an orthonormal basis. An individual eigenvector describes a mode of some linear combination of the various signal sources embedded in the covariance matrix; each eigenvector is a *subspace* of the overall signal characterization. The eigenvalue describes the power in a particular subspace. An eigenbasis can be used in a manner similar to a Fourier basis to describe the properties of a stochastic space-time

FIGURE 10-7 ■
MVDR spectrum of
eigenvectors
corresponding to the
dominant subspace,
plus one noise
subspace
eigenvector.



sequence:

$$\mathbf{x}_k = \sum_{m=1}^{NM} e_k(m) \mathbf{q}_k(m); \quad e_k(m) = \mathbf{q}_k^H(m) \mathbf{x}_k \quad (10.43)$$

The basis expansion in (10.43) follows from the Karhunen-Loève transform (KLT) [2,9]; $e_k(m)$ is the m -th Karhunen-Loève (KL) coefficient. Contrasting the KLT to the Fourier transform underlies the notion that a KL basis is data dependent, while the Fourier basis is fixed.

Figure 10-7 shows the minimum variance distortionless response (MVDR) spectrum of the first four eigenvectors of the covariance matrix leading to the PSD shown in Figure 10-5 with the addition of receiver noise with an output variance of 1 watt. The MVDR spectrum is a super-resolution view of the signal space-time characteristics; it represents the output of the space-time beamformer when using the weight vector

$$\mathbf{w}_k = \frac{\mathbf{R}_k^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)}{\mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)} \quad (10.44)$$

The beamformer output power for the space-time filter with weight vector given by (10.44) is

$$P_o(f_{sp}, \tilde{f}_d) = \frac{1}{\mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)} \quad (10.45)$$

The first three eigenvectors, corresponding to the three largest eigenvalues, point toward the three different signal sources in angle and Doppler. In general, for all eigenvectors lying in the dominant subspace (i.e., the subspace not occupied by thermal noise),

$$\mathbf{q}_k(m) \in \text{span}(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_p) \quad (10.46)$$

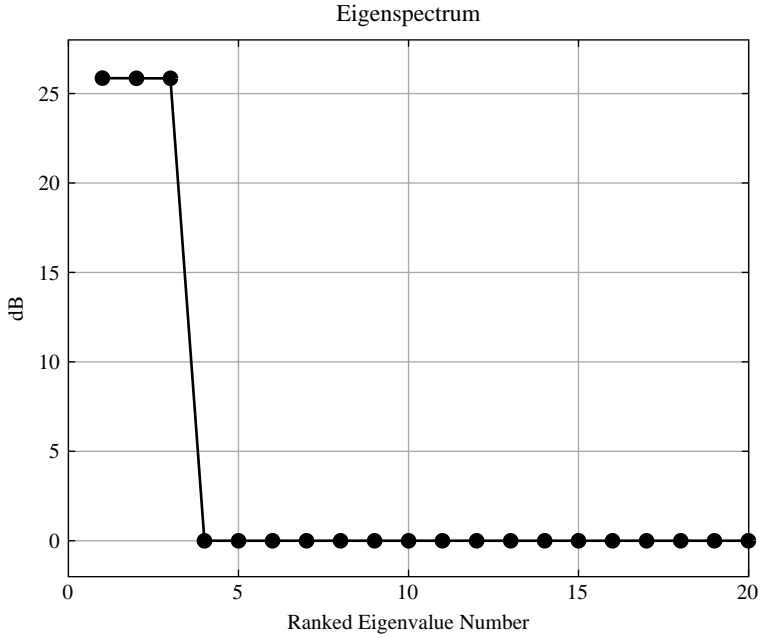


FIGURE 10-8 ■ Eigenspectrum for the three space-time signals shown in Figure 10-5, with additive receiver noise of 1 watt/channel.

where $\mathbf{s}_m \in C^{NM \times 1}$ are the various space-time signal sources $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_p$ impinging on the array. Equation (10.46) indicates each eigenvector is a linear combination of the different sources. If the sources are largely separated and of sufficient SNR, then a single source can dominate an eigensubspace, as seen in Figure 10-7.

A plot of the rank-ordered eigenvalues is known as the eigenspectrum. Identifying the colored noise subspace and relative interference-to-noise ratio (INR) or clutter-to-noise ratio (CNR) are common uses of the eigenspectra (the interference-to-noise ratio is given as the ratio of the interference and noise eigenvalues). The dominant subspace is also known as the signal or interference subspace, depending on the context, while the eigenvectors corresponding to the smaller eigenvalues comprise the noise subspace. Figure 10-8 shows the eigenspectrum for the first 20 eigenvalues of the covariance matrix used to generate Figure 10-5, with the addition of receiver noise; we find three dominant eigenvalues corresponding to the three signal sources. Since the three space-time signals are largely separated in angle and Doppler with unity amplitude weighting ($\sigma_s^2 = 1$) and the output noise is set to 1 watt ($\sigma_n^2 = 1$), the three eigenvalues of 25.85 dB are expected and correspond to the signal power times the space-time integration gain plus the noise power; that is, the eigenvalue measures power in a given subspace and is given in decibels as $10 \cdot \log_{10}(\sigma_s^2 NM + \sigma_n^2)$ when the sources are largely separated in angle and Doppler. The eigenspectrum is valuable when determining the number of DoFs required to cancel the interference: each eigenvalue, or signal source, requires one DoF for effective cancellation.

Some important properties of one-dimensional covariance matrices are given in [9–11]. Many of these properties apply to the space-time scenario. Among these various properties, it is worth mentioning two important characteristics: (1) the space-time covariance matrix is Hermitian; and (2) it is nonnegative definite and usually positive definite (as indicated by (10.41)). A Hermitian matrix equals its conjugate transpose; this apparent symmetry of the covariance matrix also applies to its inverse and is useful in

mathematical manipulations. That the space-time covariance matrix is nonnegative definite directly follows from our interpretation of the diagonal entries representing the power at each space-time element. With the addition of thermal receiver noise, the space-time covariance matrix is positive definite, signifying the guaranteed presence of its inverse. The fact that an inverse exists will prove valuable to our discussion in Section 10.5 when we develop the maximum SINR weight vector.

10.3 | SPACE-TIME PROPERTIES OF GROUND CLUTTER

In this section we leverage our discussion of space-time signals to characterize the properties of ground clutter returns. In our discussion, we model the radar return from distributed clutter as the voltage summation of individual patch returns at distinct ranges and angles. The angular location of the clutter patch with respect to the platform velocity vector determines the patch Doppler frequency; this unique coupling of spatial and Doppler frequencies yields a ridge in the clutter PSD. Since the voltage corresponding to each patch arises from the complex, coherent summation of the returns from many smaller scatterers, it is common to assume statistical independence among the various patches. Additionally, the clutter voltage is sensibly taken as uncorrelated in the range dimension. Due to the coloration of the clutter return in space and slow time, adaptive processing using these specific domains provides a mechanism for capably mitigating clutter's influence on detection performance.

The ground clutter return corresponding to the k -th range results from the coherent summation of the many scattering centers within the bounds of each iso-range, including range ambiguities. A simple yet effective model for the clutter space-time snapshot takes the form

$$\mathbf{c}_k = \sum_{m=1}^{N_c} \sum_{n=1}^{N_a} \mathbf{a}_{s-t}(m, n; k) \odot \mathbf{s}_{s-t}(f_{sp/m,n}, \tilde{f}_{d/m,n}; k) \quad (10.47)$$

where we assume each iso-range consists of N_c statistically independent clutter patches, N_a indicates the number of ambiguous ranges, $f_{sp/m,n}$ and $\tilde{f}_{d/m,n}$ represent the spatial and normalized Doppler frequencies of the $m - n$ -th patch, and $\mathbf{a}_{s-t}(m, n; k)$ is the length- NM vector containing the space-time voltages for each channel-pulse-range sample and is proportional to the square root of the clutter patch RCS. Given the platform velocity vector, $\mathbf{v}_p = v_{p,x}\hat{\mathbf{x}} + v_{p,y}\hat{\mathbf{y}} + v_{p,z}\hat{\mathbf{z}}$, the normalized clutter patch Doppler is

$$\tilde{f}_{d/m,n} = \frac{2v_{r/m,n}T}{\lambda} = \frac{2T}{\lambda} (\mathbf{v}_p \cdot \mathbf{k}(\phi_{m,n}, \theta_{m,n})) \quad (10.48)$$

$v_{r/m,n}$ is the radial, or line of sight, velocity corresponding to the $m - n$ -th clutter patch. From (10.48) we see that ground clutter Doppler has a distinct dependence on angle. Figure 10-9 pictorially characterizes the calculation in (10.47).

The clutter covariance matrix is

$$\mathbf{R}_{c/k} = E[\mathbf{c}_k \mathbf{c}_k^H] = \sum_{m=1}^{N_c} \sum_{n=1}^{N_a} \mathbf{R}_a(m, n; k) \odot \mathbf{s}_{s-t}(f_{sp/m,n}, \tilde{f}_{d/m,n}; k) \mathbf{s}_{s-t}^H(f_{sp/m,n}, \tilde{f}_{d/m,n}; k) \quad (10.49)$$

where $\mathbf{R}_a(m, n; k) = E[\mathbf{a}_{s-t}(m, n; k) \mathbf{a}_{s-t}^H(m, n; k)]$ describes the space-time voltage correlation. If the voltage is perfectly correlated from pulse-to-pulse and channel-to-channel,

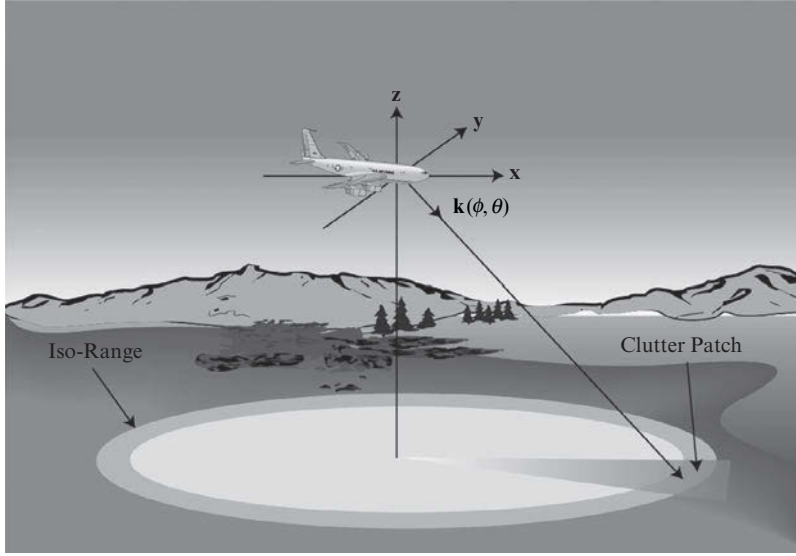


FIGURE 10-9 ■
Geometry for
space-time clutter
patch calculation
(after [8], © 2003
IEEE)

$\mathbf{R}_a(m, n; k) = \sigma_{c/m, n; k}^2 \mathbf{1}_{MN}$; $\mathbf{1}_{MN}$ is the Hadamard identity, an $NM \times NM$ matrix filled with ones, and $\sigma_{c/m, n; k}^2$ represents the clutter power for the $m - n$ -th patch at the k -th range. Under such circumstances, (10.49) simplifies to

$$\mathbf{R}_{c/k} = E[\mathbf{c}_k \mathbf{c}_k^H] = \sum_{m=1}^{N_c} \sum_{n=1}^{N_a} \sigma_{c/m, n; k}^2 \mathbf{s}_{s-t}(f_{sp/m, n}, \tilde{f}_{d/m, n}; k) \mathbf{s}_{s-t}^H(f_{sp/m, n}, \tilde{f}_{d/m, n}; k) \quad (10.50)$$

Since ground clutter at a given angle has a specific Doppler frequency dependent on the projection of the platform velocity vector onto the direction vector pointing at the clutter patch (i.e., $\mathbf{v}_p \cdot \mathbf{k}(\phi_{m, n}, \theta_{m, n})$, as seen from (10.48)), clutter resides on a ridge when viewing the angle-Doppler power spectrum [4-8]. Specifically, each clutter patch yields a sinc-like response in the angle-Doppler domain—similar to the responses shown in Figure 10-5—and thus the clutter contribution over all azimuth and at a particular iso-range traces out a ridge. In the case of a perfectly sidelooking array (SLA), and barring any angle or Doppler aliasing, the angle-Doppler clutter ridge assumes an S-shaped structure. The clutter patch at broadside exhibits a response at zero angle and zero Doppler, while points at azimuths toward the direction of flight (positive spatial frequencies) exhibit increasingly positive Doppler frequency, and points away from the direction of flight have negative spatial and Doppler frequencies. The forward-looking array (FLA) represents the extreme situation from the SLA. When considering an FLA, the clutter ridge exhibits increased curvature and symmetry, since clutter at two points on either side of the direction of flight have equal but opposite spatial frequencies yet identical Doppler frequency.

In [8], we validate the clutter model of (10.47) using measured data taken from the Multi-Channel Airborne Radar Measurements (MCARM) Program [12]. Specifically, using auxiliary variables characterizing the platform attitude and radar settings, we used (10.47) to simulate clutter data and then compared the results against the MCARM data. Overall, we show extraordinary correspondence between measured and simulated data in [8,13], thereby justifying the form given in (10.47).

FIGURE 10-10 ■
Clutter-plus-noise
PSD.

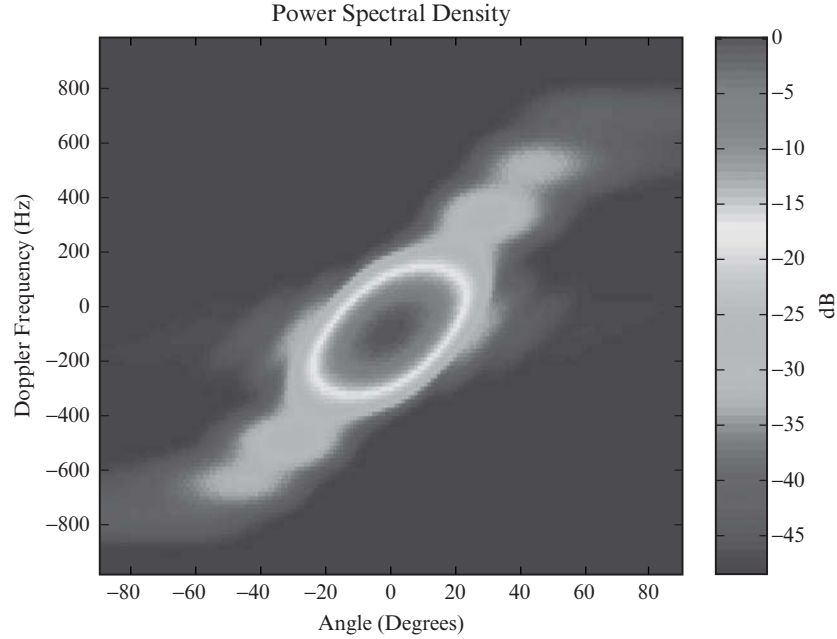


Figure 10-10 shows a typical clutter-plus-noise PSD; the clutter ridge is apparent in the figure. We simulate the data giving rise to Figure 10-10 using the MCARM parameters given in [12] with minor modification. Salient parameters include 1 GHz transmit frequency, 1 MHz waveform bandwidth, $1.33 \text{ m} \times 0.97 \text{ m}$ sidelooking receive array, six degrees of crab (yaw), 22 spatial channels ($M = 22$) in an 11-over-11 configuration, 32 pulses ($N = 32$) in the CPI, and a PRF of 1984 Hz.

The PSD in Figure 10-10 shows that the clutter power occupies only a portion of the overall two-dimensional spectrum. Additionally, the portion of the occupied spectrum exhibits a distinct angle-Doppler coupling. Detecting moving target signals competing with the clutter returns is the primary goal of the radar signal processor. Exploiting the notion that a moving target generally possesses an angle-Doppler response distinct from that of stationary clutter is the goal of the space-time filter. Some space-time filter designs yield better performance than others. In subsequent sections we first discuss generic space-time processing as a linear filtering operation applied to space-time snapshots from a given range, next consider critical MTI performance metrics, and then develop the optimal space-time filter providing the best possible detection performance for a linear filter. Consequently, we define the STAP as a data-domain implementation of the space-time optimal filter.

10.4 | SPACE-TIME PROCESSING

Figure 10-11 shows a generic space-time processing chain. After vectorizing the data matrix $\mathbf{X}_k \in C^{M \times N}$ to generate the snapshot $\mathbf{x}_k \in C^{MN \times 1}$ (i.e., $\mathbf{x}_k = \text{vec}(\mathbf{X}_k)$), the processor applies the space-time weight vector $\mathbf{w}_k \in C^{MN \times 1}$, thereby yielding the scalar output

$$y_k = \mathbf{w}_k^H \mathbf{x}_k \quad (10.51)$$

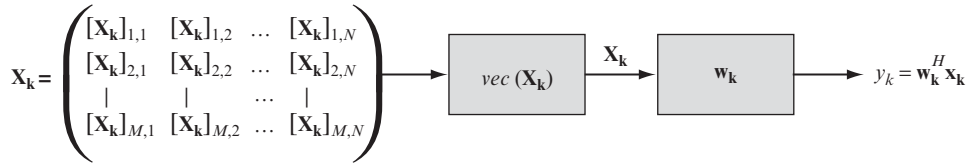


FIGURE 10-11 ■
Generic space-time
processing chain.

The processor then compares a monotonic function of this output, such as the magnitude or magnitude-squared, to a threshold to determine the presence of a target.

More specifically, radar detection generally involves binary hypothesis testing to ascertain which of two models—absence or presence of the target signal—generated the given observation [1–3,14]. The absence of target—or, equivalently, the additive combination of clutter, jamming, and noise only—is known as the null hypothesis, often written as H_0 . On the other hand, the presence of target in combination with the null hypothesis components is known as the alternative hypothesis, or H_1 . We express the two hypotheses as

$$\begin{aligned} H_0 : \mathbf{x}_{k/H_0} &= \mathbf{c}_k + \mathbf{j}_k + \mathbf{n}_k; \\ H_1 : \mathbf{x}_{k/H_1} &= \bar{\mathbf{s}}_k + \mathbf{c}_k + \mathbf{j}_k + \mathbf{n}_k = \bar{\mathbf{s}}_k + \mathbf{x}_{k/H_0} \end{aligned} \quad (10.52)$$

where $\mathbf{c}_k, \mathbf{j}_k, \mathbf{n}_k$, and $\bar{\mathbf{s}}_k$ are clutter, jamming, receiver noise, and target signal snapshots, respectively. Collectively, we refer to the presence of colored noise (clutter and jamming) and receiver noise terms as the interference-plus-noise condition. Ideally, a threshold is set so that a decision of target presence is made with sufficiently high probability when $y_k = \mathbf{w}_k^H \mathbf{x}_{k/H_1}$, while this same threshold setting leads to a threshold crossing when $y_k = \mathbf{w}_k^H \mathbf{x}_{k/H_0}$ with substantially lower probability. We define the probability of detection, P_D , as the probability of successfully choosing target presence when indeed H_1 generated the output observation; conversely, the probability of false alarm, P_{FA} , describes the probability the detector incorrectly chooses H_1 when H_0 is in fact the correct hypothesis. It is standard to set the decision threshold to achieve a specified P_{FA} ; practical considerations, such as tracking algorithm performance and onboard computing resources, determine an acceptable value of P_{FA} (nominally in the range of $P_{FA} = 10^{-6}$).

When the individual elements of \mathbf{x}_k jointly adhere to a complex normal (Gaussian) distribution, it is shown in [1–3,14] that maximizing SINR equivalently maximizes the probability of detection for a fixed probability of false alarm. The output SINR is

$$\text{SINR} = \frac{P_s}{P_{i+n}} = \frac{E[\mathbf{w}_k^H \bar{\mathbf{s}}_k \bar{\mathbf{s}}_k^H \mathbf{w}_k]}{E[\mathbf{w}_k^H \mathbf{x}_{k/H_0} \mathbf{x}_{k/H_0}^H \mathbf{w}_k]} = \frac{\mathbf{w}_k^H E[\bar{\mathbf{s}}_k \bar{\mathbf{s}}_k^H] \mathbf{w}_k}{\mathbf{w}_k^H E[\mathbf{x}_{k/H_0} \mathbf{x}_{k/H_0}^H] \mathbf{w}_k} \quad (10.53)$$

where P_s is the output signal power, and P_{i+n} is the output interference-plus-noise power. With $\bar{\mathbf{s}}_k = a_{s-t} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ and $\sigma_s^2 = E[|a_{s-t}|^2]$ representing the single-channel, single-pulse target signal power, we find

$$P_s = \mathbf{w}_k^H E[\bar{\mathbf{s}}_k \bar{\mathbf{s}}_k^H] \mathbf{w}_k = \mathbf{w}_k^H \mathbf{R}_s \mathbf{w}_k = \sigma_s^2 |\mathbf{w}_k^H \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)|^2 \quad (10.54)$$

$\mathbf{R}_s = E[\bar{\mathbf{s}}_k \bar{\mathbf{s}}_k^H] = \sigma_s^2 \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) \mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d)$ is the signal correlation matrix. We can express the interference-plus-noise power as

$$P_{i+n} = \mathbf{w}_k^H E[\mathbf{x}_{k/H_0} \mathbf{x}_{k/H_0}^H] \mathbf{w}_k = \mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k \quad (10.55)$$

We assume $\mathbf{x}_{\mathbf{k}/\mathbf{H}_0} \sim CN(\mathbf{0}, \mathbf{R}_{\mathbf{k}})$, where $\mathbf{R}_{\mathbf{k}} = E[\mathbf{x}_{\mathbf{k}/\mathbf{H}_0} \mathbf{x}_{\mathbf{k}/\mathbf{H}_0}^H]$ is the null hypothesis covariance matrix. Using (10.54)–(10.55) in (10.53) gives

$$SINR = \frac{\mathbf{w}_{\mathbf{k}}^H \mathbf{R}_{\mathbf{s}} \mathbf{w}_{\mathbf{k}}}{\mathbf{w}_{\mathbf{k}}^H \mathbf{R}_{\mathbf{k}} \mathbf{w}_{\mathbf{k}}} = \frac{\sigma_s^2 |\mathbf{w}_{\mathbf{k}}^H \mathbf{s}_{\mathbf{s}-\mathbf{t}}(f_{sp}, \tilde{f}_d)|^2}{\mathbf{w}_{\mathbf{k}}^H \mathbf{R}_{\mathbf{k}} \mathbf{w}_{\mathbf{k}}} \quad (10.56)$$

We further consider the relationship between SINR, P_D , and P_{FA} in Section 10.4.1.

Additionally, note that we have not yet specified a choice for $\mathbf{w}_{\mathbf{k}}$. Some choices include the two-dimensional matched filter (which maximizes SNR), the STAP weight vector (which attempts to maximize SINR), and the displaced phase center antenna (DPCA) weight vector (which seeks to cancel the stationary clutter signal entirely).

In the noise-limited case, $\mathbf{R}_{\mathbf{k}} \triangleq \sigma_n^2 \mathbf{I}_{NM}$. Thus, (10.56) becomes

$$SINR = \frac{\sigma_s^2 |\mathbf{w}_{\mathbf{k}}^H \mathbf{s}_{\mathbf{s}-\mathbf{t}}(f_{sp}, \tilde{f}_d)|^2}{\mathbf{w}_{\mathbf{k}}^H \mathbf{R}_{\mathbf{k}} \mathbf{w}_{\mathbf{k}}} = \frac{\sigma_s^2 |\mathbf{w}_{\mathbf{k}}^H \mathbf{s}_{\mathbf{s}-\mathbf{t}}(f_{sp}, \tilde{f}_d)|^2}{\sigma_n^2 \mathbf{w}_{\mathbf{k}}^H \mathbf{w}_{\mathbf{k}}} \quad (10.57)$$

Equation (10.57) takes the same form as (10.23), so $SINR \rightarrow SNR$. As previously discussed, the matched filter weight vector, $\mathbf{w}_{\mathbf{k}} = \mathbf{s}_{\mathbf{s}-\mathbf{t}}(f_{sp}, \tilde{f}_d)$, yields $\max(SNR) = (\sigma_s^2 / \sigma_n^2) MN$, where MN is the space-time integration gain. Additionally, colored noise always leads to degradation with respect to noise-limited performance, that is, $SNR \geq SINR$.

10.4.1 Performance Metrics

In this section we consider the following important performance metrics: probability of detection, P_D ; probability of false alarm, P_{FA} ; SINR loss; minimum detectable velocity (MDV); usable Doppler space fraction (UDSF); and improvement factor (IF).

10.4.1.1 Detection

An optimum detection statistic for the filter output $y_k = \mathbf{w}_{\mathbf{k}}^H \mathbf{x}_{\mathbf{k}}$ for the two hypotheses in (10.52), where $\mathbf{x}_{\mathbf{k}/\mathbf{H}_0} \sim CN(\mathbf{0}, \mathbf{R}_{\mathbf{k}})$, follows from the likelihood ratio test and appears as [1–3,14]

$$\begin{array}{c} H_1 \\ |y_k| > v_T \\ H_0 \end{array} \quad (10.58)$$

The performance of (10.58) is given by

$$\begin{aligned} P_{FA} &= \exp\left(\frac{-\beta_T^2}{2}\right) \\ P_D &= \int_{\beta_T}^{\infty} u \exp\left(\frac{-(u^2 + \alpha^2)}{2}\right) I_0(\alpha u) du \end{aligned} \quad (10.59)$$

where P_{FA} is the probability of false alarm, P_D is the probability of detection, β_T is a normalized detection threshold, $I_0(\cdot)$ is the modified zero-order Bessel function of the first kind, α equals the square-root of the peak output SINR, and $\beta_T = v_T / \sqrt{\mathbf{w}_{\mathbf{k}}^H \mathbf{R}_{\mathbf{k}} \mathbf{w}_{\mathbf{k}}}$ is the normalized threshold. Using (10.56), and accounting for average signal power, we

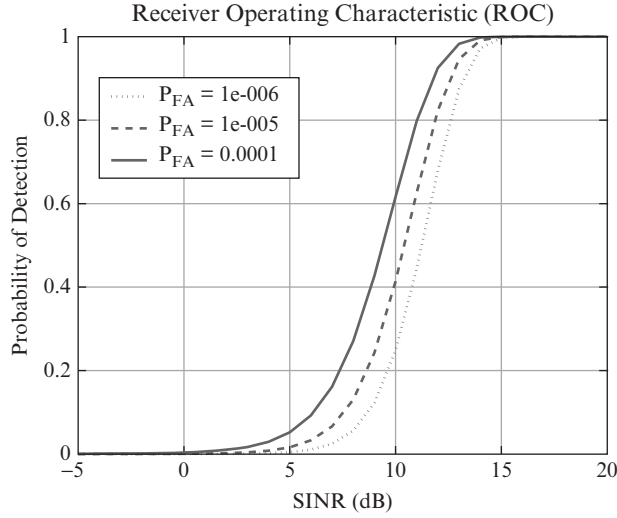


FIGURE 10-12 ■ Receiver operating characteristic corresponding to equation (10.59).

define the peak SINR as

$$\alpha^2 = 2 \times \text{SINR} = 2 \times \frac{\sigma_s^2 |\mathbf{w}_k^H \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)|^2}{\mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k} \quad (10.60)$$

Equation (10.59) is a monotonic function of α , and hence α^2 . Thus, maximizing SINR likewise maximizes P_D for a fixed value of P_{FA} .

Figure 10-12 plots P_D versus SINR (using average power, not peak power) for various values of P_{FA} . This plot is known as a receiver operating characteristic (ROC), or ROC curve. (Note: oftentimes a ROC curve plots P_D versus P_{FA} for fixed SINR, thus displaying the information in Figure 10-12 slightly different.)

10.4.1.2 SINR Loss

Using (10.1), the radar designer calculates SNR for a particular target RCS and range. Figure 10-12 is valid upon replacing SINR with SNR, as seen from (10.57). Due to the critical importance of SINR, STAP researchers commonly employ SINR loss factors to assess detection performance potential [4–8]. Two commonly used SINR loss factors are $L_{s,1}(f_{sp}, \tilde{f}_d)$ and $L_{s,2}(f_{sp}, \tilde{f}_d)$, where each loss term is bound between zero and unity. $L_{s,1}(f_{sp}, \tilde{f}_d)$ compares interference-limited performance with noise-limited capability, assuming all quantities are known:

$$L_{s,1}(f_{sp}, \tilde{f}_d) = \frac{\text{SINR}|_{\mathbf{w}_k}}{\text{SNR}} = \left(\frac{\mathbf{w}_k^H \mathbf{R}_s \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k} \right) \bigg/ \left(\frac{\sigma_s^2}{\sigma_n^2} NM \right) \quad (10.61)$$

Since the optimal weight vector calculation, discussed in Section 10.5.1, requires the known covariance matrix, $L_{s,1}(f_{sp}, \tilde{f}_d)$ is sometimes called clairvoyant SINR loss.

$L_{s,2}(f_{sp}, \tilde{f}_d)$ determines the loss between an implementation requiring estimated statistics and the clairvoyant case (e.g., adaptive versus optimum):

$$L_{s,2}(f_{sp}, \tilde{f}_d) = \frac{\text{SINR}|_{\mathbf{w}_k=\hat{\mathbf{w}}_k}}{\text{SINR}|_{\mathbf{w}_k=\mathbf{w}_{k/opt}}} = \left(\frac{\hat{\mathbf{w}}_k^H \mathbf{R}_s \hat{\mathbf{w}}_k}{\hat{\mathbf{w}}_k^H \mathbf{R}_k \hat{\mathbf{w}}_k} \right) \bigg/ \left(\frac{\mathbf{w}_{k/opt}^H \mathbf{R}_s \mathbf{w}_{k/opt}}{\mathbf{w}_{k/opt}^H \mathbf{R}_k \mathbf{w}_{k/opt}} \right) \quad (10.62)$$

Given the loss factor terms, SINR can be written

$$\text{SINR}(f_{sp}, \tilde{f}_d) = \text{SNR}(f_{sp}) \times L_{s,1}(f_{sp}, \tilde{f}_d) \times L_{s,2}(f_{sp}, \tilde{f}_d), \quad (10.63)$$

where $\text{SNR}(f_{sp})$ is the angle-dependent signal-to-noise ratio. Target velocities closest to the dominant clutter component and exhibiting SINR loss above some acceptable value, viz. $L_{s,1}(f_{sp}, \tilde{f}_d) \cdot L_{s,2}(f_{sp}, \tilde{f}_d) \geq \varepsilon$, determine the MDV. For example, suppose we calculate SNR to be 13 dB, thereby yielding $P_D = 0.87$ for $P_{FA} = 1E - 6$ according to (10.59) and Figure 10-12. If our minimum detection requirement is $P_D = 0.5$ for this same false alarm rate, then SINR must be greater than or equal to 11.25 dB. This indicates a tolerable combined SINR loss of 1.75 dB, or $\varepsilon = 0.668$.

The UDSF is that percent of the Doppler space yielding acceptable SINR loss based on detection performance requirements [5].

10.4.1.3 Improvement Factor

IF is another common metric, given as

$$\text{IF} = \frac{\text{SINR}_{out}}{\text{SINR}_{element}} = \frac{|\mathbf{w}_k^H \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)|^2 (\sigma_c^2 + \sigma_n^2)}{\mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k} \quad (10.64)$$

where σ_c^2 is the total clutter power received by a single subarray on a single pulse [4]. In the noise-limited case, (10.64) defaults to the space-time integration gain (nominally, NM). IF closely relates to the preceding SINR loss definitions.

10.5 | STAP FUNDAMENTALS

A selection for the weight vector, \mathbf{w}_k , is a key issue remaining from the discussion in the prior section. This section briefly discusses three approaches to choose the weight vector: the maximum SINR filter, the minimum variance beamformer, and the generalized sidelobe canceller. While the formulations are different, each equivalently maximizes output SINR.

10.5.1 Maximum SINR Filter

The optimal space-time weight vector maximizes the output SINR and takes the form $\mathbf{w}_k = \beta \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$, for arbitrary scalar β [3]. To see this, express (10.56) as

$$\text{SINR} = \sigma_s^2 \frac{|\hat{\mathbf{w}}_k^H \hat{\mathbf{s}}|^2}{\hat{\mathbf{w}}_k^H \hat{\mathbf{w}}_k} \leq \sigma_s^2 \frac{\hat{\mathbf{w}}_k^H \hat{\mathbf{w}}_k \hat{\mathbf{s}}^H \hat{\mathbf{s}}}{\hat{\mathbf{w}}_k^H \hat{\mathbf{w}}_k} \quad (10.65)$$

where $\mathbf{w}_k = \mathbf{R}_k^{-1/2} \hat{\mathbf{w}}_k$ and $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) = \mathbf{R}_k^{1/2} \hat{\mathbf{s}}$. For covariance matrices of interest, $\mathbf{R}_k = \mathbf{R}_k^{1/2} \mathbf{R}_k^{1/2}$. Analogous to the matched filter case, by choosing $\hat{\mathbf{w}}_k = \hat{\mathbf{s}}$ we find the left-hand side of (10.65) achieves the upper bound. Substituting the prior expressions gives

$$\hat{\mathbf{w}}_k = \hat{\mathbf{s}} \Rightarrow \mathbf{R}_k^{1/2} \mathbf{w}_k = \mathbf{R}_k^{-1/2} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) \Rightarrow \mathbf{w}_k = \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) \quad (10.66)$$

Scaling the weight vector by β does not alter the output SINR.

It is convenient to interpret the optimal space-time filtering operation in two steps: whitening and then warped matched filtering. If, for instance, we let $\mathbf{R}_k = \sigma_n^2 \mathbf{I}_{NM}$, the optimal weight vector simplifies to $\mathbf{w}_k = \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$, where we have chosen $\beta = \sigma_n^2$; we recognize the resulting filter as the matched filter. In the more general case, using the optimal weight vector in (10.51) yields

$$y_k = \mathbf{w}_k^H \mathbf{x}_k = \mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{R}_k^{-1} \mathbf{x}_k = \left(\mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{R}_k^{-1/2} \right) \left(\mathbf{R}_k^{-1/2} \mathbf{x}_k \right) = \underline{\mathbf{s}}^H \underline{\mathbf{x}}_k \quad (10.67)$$

where $\underline{\mathbf{s}} = \mathbf{R}_k^{-1/2} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ and $\underline{\mathbf{x}}_k = \mathbf{R}_k^{-1/2} \mathbf{x}_k$. The covariance matrix of $\underline{\mathbf{x}}_k$ is

$$\underline{\mathbf{R}}_k = E[\underline{\mathbf{x}}_k \underline{\mathbf{x}}_k^H] = \mathbf{R}_k^{-1/2} E[\mathbf{x}_k \mathbf{x}_k^H] \mathbf{R}_k^{-1/2} = \mathbf{R}_k^{-1/2} \mathbf{R}_k \mathbf{R}_k^{-1/2} = \mathbf{I}_{NM} \quad (10.68)$$

By assuming the form of an identity matrix, $\underline{\mathbf{R}}_k$ indicates the elements of $\underline{\mathbf{x}}_k$ are uncorrelated. White noise likewise has a covariance matrix equal to an identity matrix scaled by the single-element power, $\sigma_n^2 \mathbf{I}_{NM}$. Since $\mathbf{R}_k^{-1/2}$ whitens \mathbf{x}_k , $\mathbf{R}_k^{-1/2}$ is known as a whitening filter. The processor also applies the whitening transformation to the target signal. For this reason, the matched filter must be modified accordingly; $\underline{\mathbf{s}}$ is the corresponding warped matched filter. Substituting $\hat{\mathbf{R}}_k$ for \mathbf{R}_k and \mathbf{v}_{s-t} for $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ in the adaptive case leads to a similar interpretation.

STAP is a data domain implementation of the optimal filter with weight vector given by (10.66). In practice, both \mathbf{R}_k and $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ are unknown. The processor substitutes an estimate for each quantity to arrive at the adaptive weight vector

$$\hat{\mathbf{w}}_k = \hat{\beta} \hat{\mathbf{R}}_k^{-1} \mathbf{v}_{s-t} \quad (10.69)$$

where $\hat{\beta}$ is a scalar, \mathbf{v}_{s-t} is a surrogate for $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$, and $\hat{\mathbf{R}}_k$ is an estimate of \mathbf{R}_k . This approach is known as sample matrix inversion (SMI) [15,16]. The surrogate steering vector, \mathbf{v}_{s-t} , may differ from the exact steering vector due to slightly different steering angle and Doppler, or as a result of system errors; the argument, (f_{sp}, \tilde{f}_d) , is dropped from \mathbf{v}_{s-t} to acknowledge potential mismatch.

Some selections for $\hat{\beta}$ are more useful than others. For example, it is shown in [17,18] that

$$\hat{\beta} = 1 / \sqrt{\mathbf{v}_{s-t}^H \hat{\mathbf{R}}_k^{-1} \mathbf{v}_{s-t}} \quad (10.70)$$

yields a constant false alarm rate (CFAR) characteristic under certain operating characteristics.

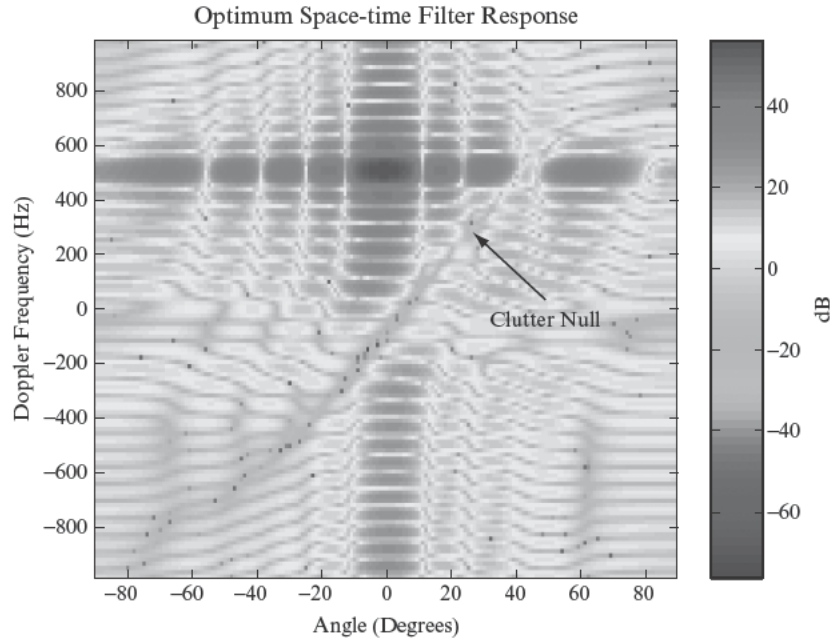
It is most common to calculate the space-time covariance matrix estimate as [15]

$$\hat{\mathbf{R}}_k = \frac{1}{P} \sum_{m=1}^P \mathbf{x}_m \mathbf{x}_m^H \quad (10.71)$$

$\{\mathbf{x}_m\}_{m=1}^P$ are known as secondary or training data. If the training data are independent and identically distributed (IID) and there is no steering vector mismatch (i.e., $\mathbf{v}_{s-t} = \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$), Reed, Mallett, and Brennan (RMB) showed $L_{s,2}(f_{sp}, \tilde{f}_d)$ is beta-distributed with mean [15]

$$E[L_{s,2}(f_{sp}, \tilde{f}_d)] = \frac{(P+2-NM)}{(P+1)} \quad (10.72)$$

FIGURE 10-13 ■
Optimum filter response over angle and Doppler, with filter tuned to broadside and 500 Hz Doppler.



P is the number of training data vectors. Equation (10.72) suggests a nominal training requirement of $P \approx 2NM$ training data vectors to achieve an average loss of 3 dB between adaptive and optimal filters; this result is known as the RMB rule after its originators. To avoid target self-whitening, the processor commonly excludes the cell under test as well as several adjacent guard cells from the training data set.

STAP provides two-dimensional filtering of the space-time data. The STAP filter response is given by

$$H(f_{sp}, \tilde{f}_d) = \mathbf{w}_k^H \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) \quad (10.73)$$

Figure 10-13 shows an example optimal magnitude frequency response for the clutter-plus-noise signal environment given by the PSD in Figure 10-10; we steer the filter response to zero degrees azimuth and elevation and a Doppler frequency of 500 Hz. The STAP response converges to this filter characteristic. This response arises by stepping $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ over all spatial and Doppler frequencies of interest. When the sampling is uniform in space and time, and the space-time aperture is error-free, we can reformat \mathbf{w}_k into a two-dimensional space-time matrix—in the same form as \mathbf{X}_k given by (10.35)—and then compute the filter response via a two-dimensional Fourier transform, viz. $\mathfrak{F}_2\{\mathbf{w}_k\}$.

Using the optimal weight vector of (10.66) in (10.53) yields an expression for the maximum SINR

$$\text{SINR}_{\max} = \sigma_s^2 \mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) \quad (10.74)$$

SINR loss is a more convenient metric, since it removes the influence of σ_s^2 ; inserting the optimal weight vector into the SINR loss expression of (10.61) provides a bound on performance for all linear, space-time processing methods. The clairvoyant (known

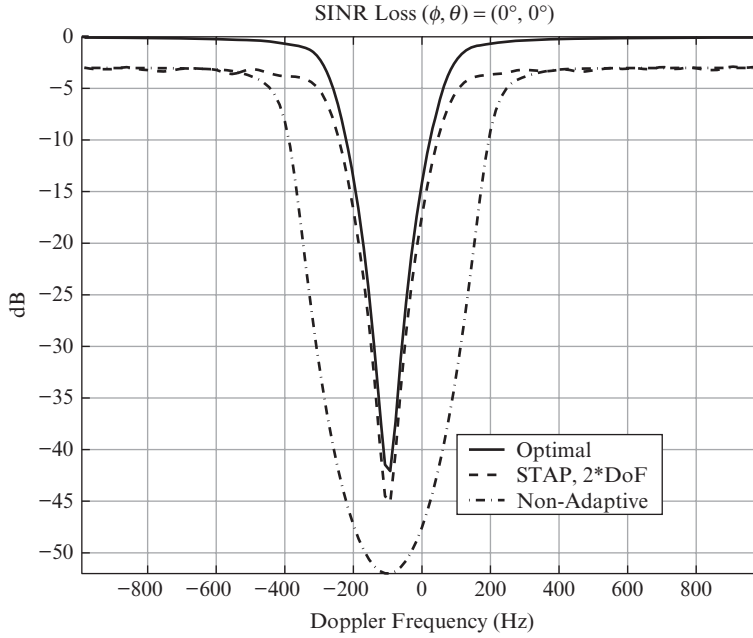


FIGURE 10-14 ■ SINR loss comparison for optimal filter, STAP, and nonadaptive processing approaches.

covariance matrix) SINR loss takes the form

$$L_{s,1/opt}(f_{sp}, \tilde{f}_d) = \frac{\mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{R}_{\mathbf{k}}^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)}{NM/\sigma_n^2} \quad (10.75)$$

Figure 10-14 shows SINR loss corresponding to the data previously used to generate the results given in Figure 10-10 and Figure 10-13. SINR loss curves are given for the optimal space-time processor; STAP employing an IID training set of size $2 * \text{DoF} = 2NM$, or 1408 training bins at 150 m range resolution; and a nonadaptive beamformer implementation with a Hanning weight in Doppler, a 25 dB Taylor weighting in azimuth, and no weighting in elevation. We note from this figure the following key points: 0 dB represents noise-limited performance; the STAP nominally suffers a 3 dB loss with respect to optimal as a result of the finite extent of the training data; and the performance of the STAP far exceeds that of the nonadaptive space-time beamformer.

10.5.2 Minimum Variance Beamformer

The MV space-time beamformer is another common formulation for the adaptive processor and employs a weight vector yielding minimal output power subject to a linear constraint on the desired target spatial and temporal response:

$$\min_{\mathbf{w}_{\mathbf{k}}} E[y_k y_k^*] \text{ subject to } \mathbf{w}_{\mathbf{k}}^H \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) = g \quad (10.76)$$

g is a complex scalar. The minimum variance weight vector is [2,9]

$$\mathbf{w}_{\mathbf{k}} = \frac{g^* \mathbf{R}_{\mathbf{k}}^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)}{\mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{R}_{\mathbf{k}}^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)} \quad (10.77)$$

Setting $g = 1$ is known as the distortionless response and leads to the weight vector in (10.44) [9]. An expression for the output power of this beamformer is given by (10.45).

We find that (10.77) is of the form $\mathbf{w}_k = \beta \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$, where

$$\beta = \frac{g^*}{\mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)} \quad (10.78)$$

Hence, (10.77) likewise yields maximal SINR.

As in Section 10.5.1, in practice the adaptive processor substitutes \mathbf{v}_{s-t} as a surrogate for $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ and $\hat{\mathbf{R}}_k$ as an estimate of \mathbf{R}_k .

10.5.3 Generalized Sidelobe Canceller

The GSC is a formulation that conveniently converts the minimum variance constrained beamformer described in the preceding section into an unconstrained form [9,19]. Some prefer the GSC interpretation of STAP over the whitening filter-warped matched filter interpretation of Section 10.5.1; since the GSC implements the MV beamformer for the single linear constraint [19], this structure likewise can be shown to maximize output SINR. However, this approach has computational drawbacks: as the reader will see, the interference-plus-noise covariance matrix is recomputed every time the target steering vector—and, hence, the blocking matrix—changes.

Figure 10-15 provides a block diagram of the GSC. The top leg of the GSC provides a quiescent response by forming a beam at the angle and Doppler of interest. A blocking matrix, $\mathbf{B}_G = \text{null}(\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d))$, prevents the target signal from entering the lower leg; essentially, the blocking matrix forms a notch filter tuned to (f_{sp}, \tilde{f}_d) . With the desired signal absent from the lower leg, the processor tunes the weight vector to provide a minimal mean square error (MMSE) estimate of the interference in the top leg. In the final step, the processor differences the desired signal, $d_{o/k}$, with the estimate from the lower leg, $\hat{d}_{o/k}$, to form the filter output, $y_k = d_{o/k} - \hat{d}_{o/k}$. Ideally, any large residual at the output corresponds to an actual target.

The desired signal is given as

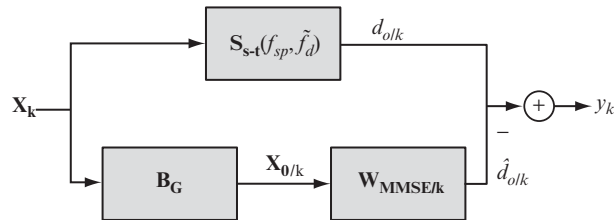
$$d_{o/k} = \mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{x}_k \quad (10.79)$$

The signal vector in the lower leg is

$$\mathbf{x}_{0/k} = \mathbf{B}_G \mathbf{x}_k; \quad \mathbf{B}_G \in C^{(NM-1) \times NM}; \quad \mathbf{x}_{0/k} \in C^{(NM-1) \times 1} \quad (10.80)$$

Forming the quiescent response of (10.79) uses a single DoF, resulting in the reduced dimensionality of $\mathbf{x}_{0/k}$. By weighting the signal vector in the lower leg, the GSC forms a

FIGURE 10-15 ■
GSC block diagram.



MMSE estimate of $d_{o/k}$ as

$$\hat{d}_{o/k} = \mathbf{w}_{\text{MMSE}/k}^H \mathbf{x}_{0/k} \quad (10.81)$$

where the MMSE weight vector follows from the well-known Wiener-Hopf equation [2,9]

$$\mathbf{w}_{\text{MMSE}/k} = \mathbf{R}_{\mathbf{x}_{0/k}}^{-1} \mathbf{r}_{\mathbf{x}_{0/k}d} \quad (10.82)$$

The lower leg covariance matrix is

$$\mathbf{R}_{\mathbf{x}_{0/k}} = E[\mathbf{x}_{0/k} \mathbf{x}_{0/k}^H] = \mathbf{B}_G \mathbf{R}_k \mathbf{B}_G^H \quad (10.83)$$

while the cross-correlation between lower and upper legs is

$$\mathbf{r}_{\mathbf{x}_{0/k}d} = E[\mathbf{x}_{0/k} d_{o/k}^*] = \mathbf{B}_G \mathbf{R}_k \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) \quad (10.84)$$

The GSC filter output is then

$$\begin{aligned} y_k &= d_{o/k} - \hat{d}_{o/k} = \mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{x}_k - \mathbf{w}_{\text{MMSE}/k}^H \mathbf{x}_{0/k} \\ &= \left(\mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) - \mathbf{w}_{\text{MMSE}/k}^H \mathbf{B}_G \right) \mathbf{x}_k. \end{aligned} \quad (10.85)$$

From (10.85) and prior discussion, it is seen that the corresponding space-time weight vector is

$$\mathbf{w}_{\text{GSLC}} = \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) - \mathbf{B}_G^H \mathbf{w}_{\text{MMSE}/k} \quad (10.86)$$

We compute the output SINR as the ratio of output signal power to interference-plus-noise power, as in (10.56). The signal-only output power of the GSC is

$$\begin{aligned} P_s &= E[\mathbf{w}_{\text{GSLC}}^H \tilde{\mathbf{s}}_k \tilde{\mathbf{s}}_k^H \mathbf{w}_{\text{GSLC}}] = E[\mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \tilde{\mathbf{s}}_k \tilde{\mathbf{s}}_k^H \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)] \\ &= \sigma_s^2 |\mathbf{s}_{s-t}^H(f_{sp}, \tilde{f}_d) \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)|^2 = \sigma_s^2 (NM)^2 \end{aligned} \quad (10.87)$$

and the output interference-plus-noise power is

$$P_{i+n} = E[\mathbf{w}_{\text{GSLC}}^H \mathbf{x}_{k/H_0} \mathbf{x}_{k/H_0}^H \mathbf{w}_{\text{GSLC}}] = \mathbf{w}_{\text{GSLC}}^H \mathbf{R}_k \mathbf{w}_{\text{GSLC}} \quad (10.88)$$

In practice, just as for the maximum SINR filter and the MV beamformer, the adaptive processor substitutes \mathbf{v}_{s-t} as a surrogate for $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ and $\hat{\mathbf{R}}_k$ as an estimate of \mathbf{R}_k .

The multistage Wiener filter exploits the GSC structure, providing an iterative, signal-dependent, rank reduction strategy that enhances adaptive filter convergence [20].

10.6 | STAP PROCESSING ARCHITECTURES AND METHODS

Section 10.5 discussed several approaches to optimally and adaptively weight a space-time data vector, \mathbf{x}_k . The resulting space-time weighting invariably involves a steering vector matched to the target response over the space-time aperture and an inverse covariance matrix to mitigate colored interference, such as clutter.

Two principal challenges when implementing the space-time processor include computational loading, which is roughly cubic in the DoFs, and the limited availability of suitable training data (homogeneous training vectors of quantity at least twice the processor's DoFs) to estimate the unknown interference-plus-noise covariance matrix. RD-STAP methods given in this section address these challenges and are of practical significance. This section also briefly discusses reduced-rank STAP (RR-STAP) [21,22], which is a weight calculation method and can be used with the space-time data or RD-STAP methods. An integrated approach to estimate target angle and Doppler is also given, along with an overview of an end-to-end processing block diagram.

10.6.1 Reduced-Dimension STAP

In the previous sections we described STAP as a two-dimensional, adaptive, linear filter operating on M spatial channels and N pulses. This direct formulation is known as the joint domain STAP. Critical joint domain STAP limitations include a need for substantial training sample support and high computational burden. In accord with the RMB rule, nominal training support is approximately $2NM$; since typically $32 \leq N \leq 128$ and $6 \leq M \leq 10$, training data can easily cover 384 to 2560 bins (for comparison, a typical CFAR window covers from 30 to 40 bins to avoid variable clutter features). Additionally, computational burden associated with the SMI approach is $O(N^3 M^3)$. To overcome these limitations without affecting the space-time aperture (i.e., without reducing either N or M , thereby affecting the overall coherent gain), researchers have developed a variety of STAP techniques based on reducing the processor's dimensionality without substantially sacrificing performance. In this section of the paper we specifically focus on introducing the reader to RD-STAP methods [4,5,23–25].

In RD-STAP, typically a linear, frequency domain transformation projects the space-time data vector \mathbf{x}_k into a lower dimensional subspace. The transformed data vector is

$$\tilde{\mathbf{x}}_k = \mathbf{T}^H \mathbf{x}_k; \mathbf{T} \in C^{NM \times J} \quad (10.89)$$

where $J \ll NM$ and $\tilde{\mathbf{x}}_k$ has dimension $J \times 1$. Computational burden associated with matrix inversion drops from $O(N^3 M^3)$ to $O(J^3)$, and nominal sample support decreases from $2NM$ to $2J$. For example, the transformation may spatially or temporally beamform the data; subsequently, taking advantage of the compressed nature of the clutter data in the frequency domain (see Figure 10-10), the RD-STAP then selects several adjacent frequency bins near the target angle and Doppler of interest as one approach to implement the adaptive canceller. In other words, in RD-STAP, the processor aims to cancel clutter signals in the vicinity of the target space-time response to aggregate the largest detection performance benefit. We will consider two illustrative example architectures momentarily.

The $J \times J$ null-hypothesis covariance matrix corresponding to (10.89) is

$$\tilde{\mathbf{R}}_k = E \left[\tilde{\mathbf{x}}_{k/H_0} \tilde{\mathbf{x}}_{k/H_0}^H \right] = \mathbf{T}^H \mathbf{R}_k \mathbf{T} \quad (10.90)$$

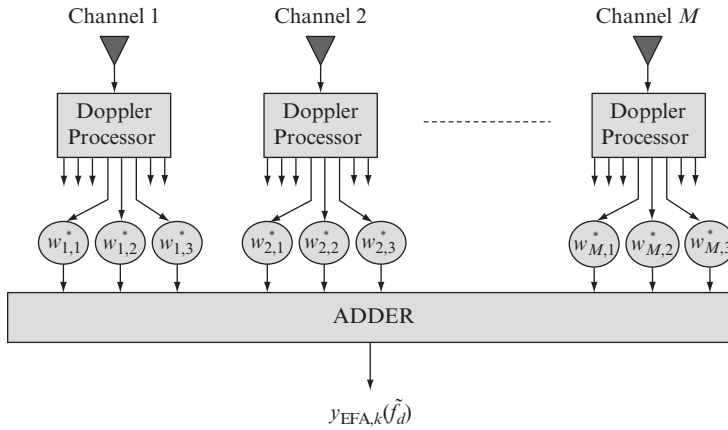
Applying the same transformation to the space-time steering vector gives $\tilde{\mathbf{s}} = \mathbf{T}^H \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$. The corresponding optimal weight vector is $\tilde{\mathbf{w}}_k = \tilde{\beta} \tilde{\mathbf{R}}_k^{-1} \tilde{\mathbf{s}}$, for arbitrary scalar $\tilde{\beta}$. The adaptive solution involves calculating $\hat{\mathbf{R}}_k$ from (10.71) using the transformed training data set $\{\mathbf{T}^H \mathbf{x}_m\}_{m=1}^P$, replacing $\tilde{\mathbf{s}}$ with the hypothesized steering vector $\tilde{\mathbf{v}}$, and

then forming and applying the adaptive weight vector $\hat{\mathbf{w}}_{\mathbf{k}} = \hat{\beta} \hat{\mathbf{R}}_{\mathbf{k}}^{-1} \hat{\mathbf{v}}$, with $\hat{\beta}$ a scalar. The RD-STAP output is $\tilde{y}_k = \hat{\mathbf{w}}_{\mathbf{k}}^H \tilde{\mathbf{x}}_{\mathbf{k}}$.

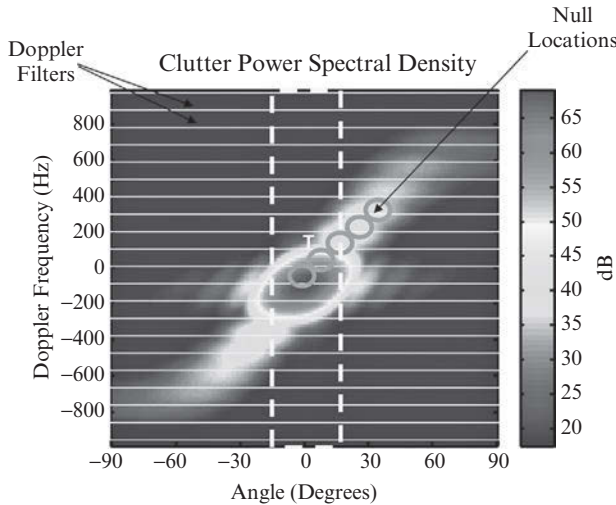
10.6.1.1 Element Space Post-Doppler

Perhaps the most popular RD-STAP method is the extended factored algorithm (EFA) given in [23]. This approach, shown in Figure 10-16a, involves Doppler filtering each receive channel and then adaptively combining several adjacent Doppler filters and all spatial channels. For example, three to five Doppler DoFs is common, so the overall space-time DoFs is $3M$ to $5M$. The adaptive weights are shown in the figure as $w_{m,n}$, where m indicates channel number, and n characterizes the relative Doppler bin number (from the three bins of interest); the superscript, $*$, denotes conjugation; and $y_{EFA,k}(\tilde{f}_d)$ is the EFA output for normalized Doppler, \tilde{f}_d .

Figure 10-16b describes EFA operation in the angle-Doppler domain. The horizontal, rectangular boxes identify various Doppler filters. EFA is able to configure nulls along a section of the angle-Doppler clutter ridge spanned by the reduced set of Doppler filter



(a) EFA RD-STAP processing approach



(b) Illustration of EFA operation in the angle-Doppler domain

FIGURE 10-16 ■
Extended factored
algorithm
reduced-dimension
STAP.

outputs, shown as circles in the figure. (This example shows five adjacent Doppler filters being combined.) The local placement of the angle-Doppler filter nulls effectively suppresses clutter in the vicinity of the target, denoted by the letter T in the figure.

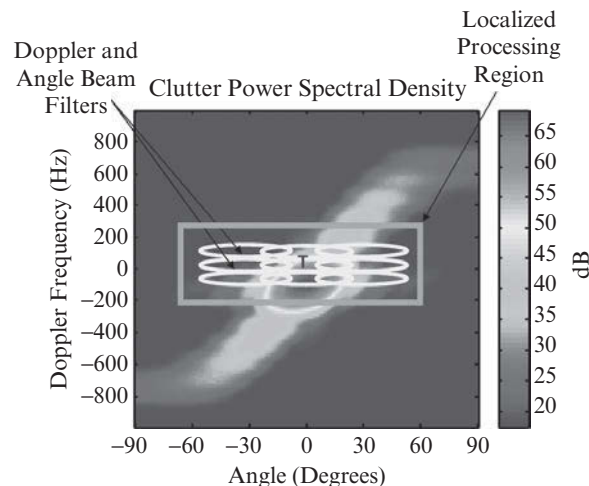
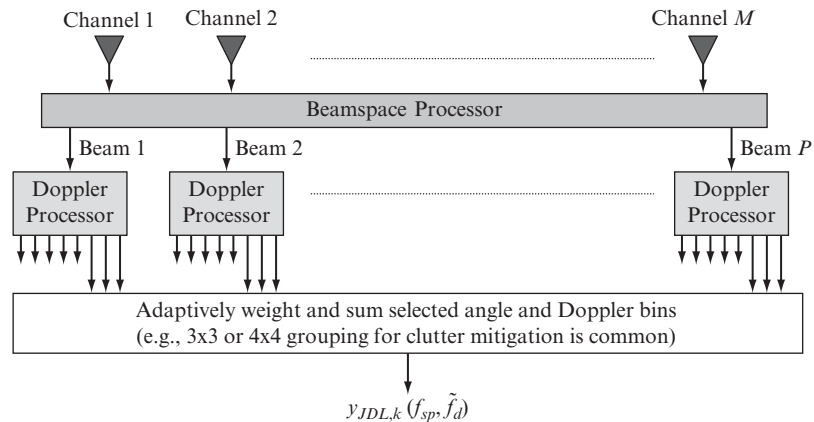
EFA offers performance close to the bound set by the joint domain optimum solution but requires a fraction of the training data ($6M$ to $10M$ versus NM) and significantly reduces computational burden. Additionally, since practical radar environments appear heterogeneous, or non-IID, for a variety of reasons [26–28], the minimal training requirements of the EFA algorithm can lead to performance exceeding the joint domain SMI implementation.

As a final note, EFA is often called element-space, post-Doppler, reduced-dimension STAP [4,5].

10.6.1.2 Post-Doppler Beamspace

The joint domain localized (JDL) algorithm is another popular STAP method [24]. Figure 10-17a shows a block diagram of this approach, which involves three steps: (1) beamforming the available spatial channels; (2) Doppler filtering each spatial beam;

FIGURE 10-17 ■
JDL reduced-dimension STAP.



and (3) adaptively combining several adjacent angle and Doppler bins about the target angle and Doppler, yielding the output $y_{JDL,k}(f_{sp}, \hat{f}_d)$. The collections of angle-Doppler bins comprise the local processing regions, with each generally centering on potential target bin locations of interest. Typical implementations employ three to five adaptive spatial beams by three to five adaptive Doppler filters [24]. The adaptive weights, $w_{m,n}$, (not shown in the figure) combine the P beams and three Doppler filters.

Figure 10-17b illustrates JDL operation. In essence, JDL works like a two-dimensional sidelobe canceller: the processor weights angle-Doppler beams surrounding the target beam (denoted with a T in the figure) as a means of mitigating clutter competing with the target. This approach localizes the selection of DoFs to best represent and suppress clutter competing with the target at a nearby angle-Doppler location.

JDL offers a diversity of spatial and temporal DoFs distinct from EFA; the benefits of both reduced training sample support (e.g., $3P$ DoFs for three Doppler bin JDL implementation, versus NM) and computational burden are tremendous. As in the EFA case, JDL can offer performance close to the bound set by the joint domain filter, but JDL is a bit more sensitive to implementation choices, such as the spacing of the spatial beams.

It is apparent from this discussion that JDL is a beamspace, post-Doppler, reduced-dimension STAP technique and is often called such in other sources of literature [5].

10.6.1.3 Post-Doppler STAP Performance

Figure 10-18 compares the performance of EFA and JDL to the optimal and SMI bounds shown in Figure 10-14; the same example and synthetic data leading to Figure 10-14 is used in this case. Both EFA and JDL implementations use a Hanning Doppler weighting, no spatial weighting, and training sample support equal to twice the processor's DoFs. Additionally, the EFA implementation uses all 22 spatial DoFs and three adjacent Doppler bins, while we apply the JDL method using three adjacent Doppler bins and three adjacent

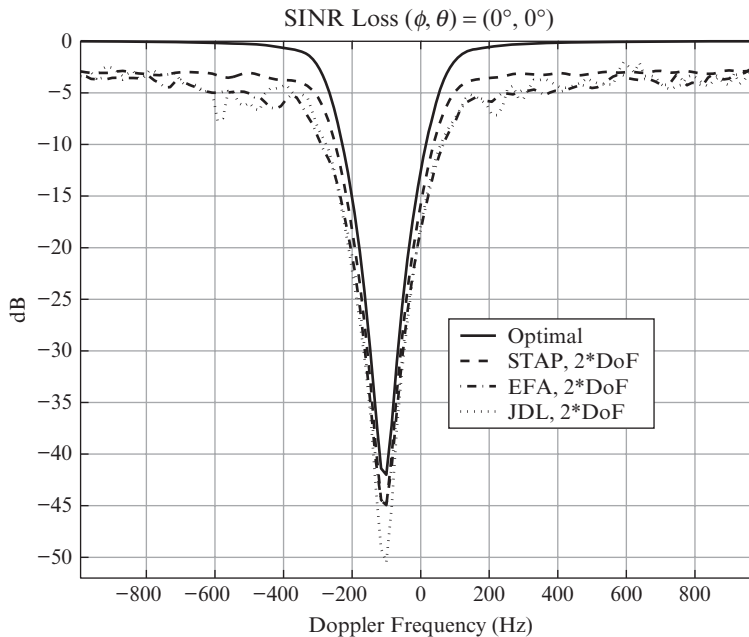


FIGURE 10-18 ■ SINR loss comparison for joint domain optimal and STAP approaches, along with performance of EFA and JDL RD-STAP methods.

spatial beams. Hence, the EFA method has 66 adaptive DoFs available for clutter suppression and JDL uses 9 DoFs, respectively. We see that both JDL and EFA perform close to the bound on adaptive performance set by the joint domain STAP case. However, note that the joint domain STAP trains over 1408 range bins (which may not be available in practice or may be corrupted by clutter heterogeneity [26–28]), compared with 133 training bins for EFA and 18 bins for JDL. Computational burden is also much lower for EFA and JDL implementations.

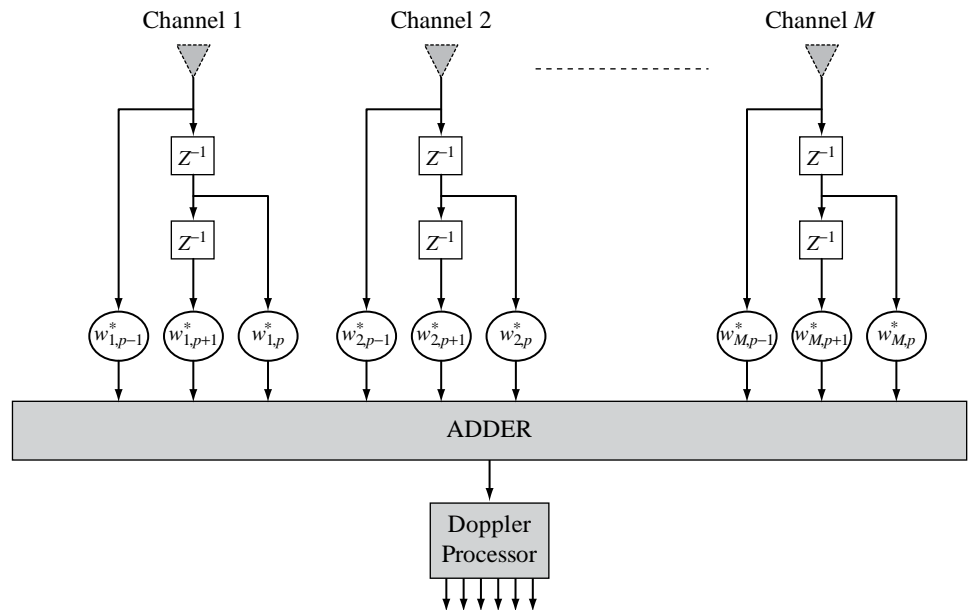
10.6.2 Pre-Doppler STAP

Details of pre-Doppler STAP are given in [4,25]; the approach is sometimes called adaptive DPCA. Figure 10-19 indicates the processing strategy. The basic idea behind pre-Doppler STAP is to take overlapping sub-CPIs composed of \tilde{N} pulses ($\tilde{N} = 3$ is a common choice) for all M channels; whiten the sub-CPI data using an $\tilde{N}M \times \tilde{N}M$ covariance matrix (which may vary among sub-CPIs); apply the Kronecker product of a temporal steering vector composed of the binomial weights and the spatial steering vector for the desired receive angle to each sub-CPI, where the binomial weights further mitigate stationary clutter while passing signals at higher Doppler frequencies to subsequent processing; and then Doppler process the clutter cancelled pulses aggregated from the output of each processed sub-CPI. If, for example, $\tilde{N} = 3$, there will be $N - \tilde{N} + 1$ sub-CPIs and consequently the same number of data points sent to the Doppler processing stage. Specifically, the ADPCA weight vector is

$$\hat{\mathbf{w}}_{\text{adpca}/k} = \hat{\mathbf{R}}_{\text{adpca}/k}^{-1} \left(\left(\left(\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 1 \\ \exp(j2\pi f_1) \\ \exp(j2\pi 2f_1) \end{bmatrix} \right) \otimes \mathbf{v}_s(f_{sp}) \right) \right) \quad (10.91)$$

where f_1 is the peak clutter Doppler times the PRI, and the binomial weights are identified as the first column vector composing the temporal steering vector, $[1 \ -2 \ 1]^T$. The

FIGURE 10-19 ■
ADPCA reduced-
dimension STAP.



covariance matrix estimate, $\hat{\mathbf{R}}_{\text{adpca}/k}$, corresponds to the data vector for M channels and \tilde{N} pulses, for example, $[\mathbf{x}_{k/s}^T(n-1) \quad \mathbf{x}_{k/s}^T(n) \quad \mathbf{x}_{k/s}^T(n+1)]^T$ for $\tilde{N} = 3$.

Pre-Doppler STAP is not as popular as post-Doppler STAP, since its performance does not generally benchmark as well relative to the joint domain optimal solution. This approach is primarily of interest when the clutter within the CPI is expected to mildly decorrelate. In such cases, pre-Doppler STAP typically adjusts the adaptive weights more frequently in time than a post-Doppler STAP method, which may prove beneficial.

10.6.3 Reduced-Rank STAP

RR-STAP is a weight calculation strategy involving an eigendecomposition of the interference-plus-noise covariance matrix and then selection of the dominant interference subspace to essentially form an orthogonal projection matrix [21,22]. The orthogonal projection matrix coherently removes the clutter signal prior to the application of a matched filter. RR-STAP is a viable strategy because the clutter signal tends to be of low numerical rank. RR-STAP methods are computationally demanding due to the required eigendecomposition but generally show improved convergence over SMI, with training vector support reducing from twice the processor's DoFs (e.g., $2NM$) to twice the interference rank for an average adaptive loss of 3 dB.

It is shown in [21] that the STAP weight vector of (10.66), with the arbitrary scaling, β , can be written

$$\mathbf{w}_k = \beta \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) = \frac{\beta}{\lambda_0} \left[\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) - \sum_{m=1}^{NM} \frac{\lambda_k(m) - \lambda_0}{\lambda_k(m)} \vartheta_k(m) \mathbf{q}_k(m) \right] \quad (10.92)$$

where $\lambda_0 = \min(\{\lambda_k(m)\}_{\forall m})$ is the noise-floor eigenvalue level, and $\vartheta_k(m) = \mathbf{q}_k^H(m) \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ is the projection of the m -th eigenvector onto the quiescent pattern. As seen from (10.92), the STAP response appears as a notching of the space-time beampattern given by $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ by the weighted, interference eigenvectors. When $\lambda_k(m) = \lambda_0$, no subtraction occurs, since the corresponding eigenvector lies in the noise subspace. Thus, running the sum from $m = 1$ to $m = r_c$, where $r_c \ll NM$ is the clutter rank, yields a more robust adaptive response, with better sidelobe pattern behavior and more rapid convergence; this implementation of (10.92) is called reduced rank.

A special case of (10.92) occurs when $\lambda_k(m) \gg \lambda_0$ for all $m = 1 : r_c$. Under such circumstances, we have from (10.92)

$$\mathbf{w}_k \approx \frac{\beta}{\lambda_0} \left[\mathbf{I}_{NM} - \sum_{m=1}^{r_c} \mathbf{q}_k(m) \mathbf{q}_k^H(m) \right] \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) \quad (10.93)$$

The term in brackets in (10.93) is an orthogonal projection; it removes all subspaces aligning with the $\mathbf{q}_k(m)$ defining the interference subspace. Naturally, after applying the orthogonal projection, $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ beamforms the data in space and time. This approach to reduced-rank processing is called principal components inverse (PCI) [22].

As noted, RR-STAP is a weight calculation strategy. Equations (10.92) and (10.93) can be used with RD-STAP methods if desired.

10.6.4 Angle and Doppler Estimation

A maximum likelihood estimate for target bearing and Doppler integrates seamlessly with STAP clutter mitigation steps. The approach described here follows [29].

Consider the alternative hypothesis of (10.52)

$$\mathbf{x}_k = \bar{\mathbf{s}}_k + \mathbf{x}_{k/H_0} = \alpha \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d) + \mathbf{x}_{k/H_0} \quad (10.94)$$

where $\bar{\mathbf{s}}_k = \alpha \mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$ is the target space-time response, α is a complex constant, and $\mathbf{x}_{k/H_0} = \mathbf{c}_k + \mathbf{j}_k + \mathbf{n}_k$ is the interference-plus-noise (total noise) vector. The objective is to estimate $\mathbf{p} = [f_{sp} \ \tilde{f}_d]^T$ with $\hat{\mathbf{p}}$. Since $\mathbf{x}_{k/H_0} \sim CN(\mathbf{0}, \mathbf{R}_k)$, the joint probability density function is

$$p(\mathbf{x}_k|\mathbf{p}) = \frac{1}{\pi^M |\mathbf{R}_k|} \exp \left(-(\mathbf{x}_k - \alpha \mathbf{s}_{s-t}(\mathbf{p}))^H \mathbf{R}_k^{-1} (\mathbf{x}_k - \alpha \mathbf{s}_{s-t}(\mathbf{p})) \right) \quad (10.95)$$

Given the likelihood function defined by (10.95), we first require an estimate for the complex constant. Equation (10.95) is maximal when

$$Q(\alpha, \mathbf{p}) = (\mathbf{x}_k - \alpha \mathbf{s}_{s-t}(\mathbf{p}))^H \mathbf{R}_k^{-1} (\mathbf{x}_k - \alpha \mathbf{s}_{s-t}(\mathbf{p})) \quad (10.96)$$

is minimal. Differentiating (10.96) and setting the result to zero yields

$$\hat{\alpha} = \frac{\mathbf{s}_{s-t}^H(\mathbf{p}) \mathbf{R}_k^{-1} \mathbf{x}_k}{\mathbf{s}_{s-t}^H(\mathbf{p}) \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(\mathbf{p})} \quad (10.97)$$

Substituting (10.97) into (10.96) leads to

$$Q(\hat{\alpha}, \mathbf{p}) = \mathbf{x}_k^H \mathbf{R}_k^{-1} \mathbf{x}_k - \frac{|\mathbf{x}_k^H \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(\mathbf{p})|^2}{\mathbf{s}_{s-t}^H(\mathbf{p}) \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(\mathbf{p})} \quad (10.98)$$

Differentiating (10.98) with respect to \mathbf{p} and setting the result to zero yields the maximum likelihood estimate, $\hat{\mathbf{p}}$. Consequently, the estimator is [29]

$$\hat{\mathbf{p}} = \begin{bmatrix} \hat{f}_{sp} \\ \hat{\tilde{f}}_d \end{bmatrix} = \arg \max_{\mathbf{p} = [f_{sp} \ \tilde{f}_d]^T} \left(\frac{|\mathbf{s}_{s-t}^H(\mathbf{p}) \mathbf{R}_k^{-1} \mathbf{x}_k|^2}{\mathbf{s}_{s-t}^H(\mathbf{p}) \mathbf{R}_k^{-1} \mathbf{s}_{s-t}(\mathbf{p})} \right) \quad (10.99)$$

The maximum likelihood estimate takes the form of an adaptive matched filter (AMF) normalized [17,18] STAP cost surface where in practice, $\mathbf{v}_{s-t}(f_{sp}, \tilde{f}_d)$ replaces $\mathbf{s}_{s-t}(f_{sp}, \tilde{f}_d)$, and $\hat{\mathbf{R}}_k$ replaces \mathbf{R}_k . Implementing the estimator requires a very fine grid search to find the peak of the likelihood function and occurs after the STAP detects a target.

10.6.5 Processing Block Diagram

Figure 10-20 shows a radar detection block diagram incorporating STAP, post-detection integration (PDI), CFAR thresholding, and bearing and Doppler estimation as the primary signal processing operations.

The multichannel array receives multiple pulses of data for range cell k , as described in preceding sections. After downconversion to an intermediate frequency and ADC, the digital processor converts the data to complex baseband and performs pulse compression

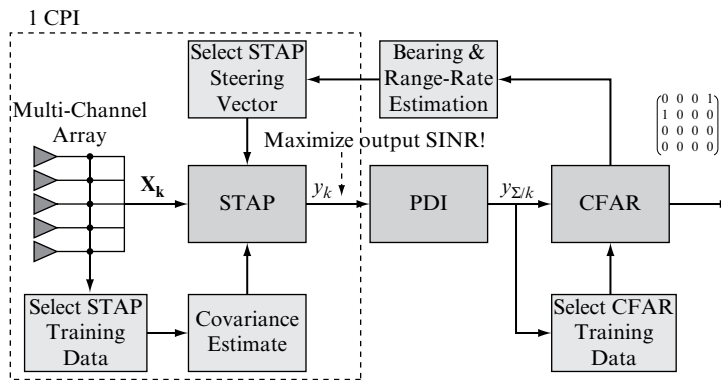


FIGURE 10-20 ■
Radar detection
block diagram
incorporating STAP.

to yield the space-time snapshot, \mathbf{x}_k . Subsequent processing intends to determine whether target energy is present in \mathbf{x}_k .

The STAP operates on \mathbf{x}_k as previously described, suppressing ground clutter and possibly jamming signals. Generally, STAP generates a response for all unambiguous Doppler frequencies at a single receive angle. The processor often implements an AMF-normalized RD-STAP method to mitigate training requirements and computational burden and to ensure residual clutter, jamming, and noise are properly referenced. In essence, the processor forms a range/range-rate map for a receive angle aligning with the expected target return; range-rate follows from Doppler frequency through multiplication by half a wavelength.

It is common that the transmitter frequency hop to avoid target fading; each CPI consists of M channels, N pulses, and L range bins collected at a common frequency. Thus, if the transmitter employs F different frequencies, each frequency yields a CPI (and corresponding data cube) to which the processor applies STAP, resulting in F range/range-rate maps. The processor ensures the pixels of the F different range/range-rate maps align. Since the AMF normalization is most typically chosen, the pixels comprising the range/range-rate map already correspond to the magnitude squared of the STAP output. In this case, PDI simply involves summing the range/range-rate pixels. A CFAR algorithm is used to estimate the ambient noise power and set a detection threshold.

Given a detection, the radar processor then estimates the target bearing, or angular position, and the target range-rate or Doppler frequency as well. The maximum likelihood estimation procedure described in Section 10.6.4 is modified slightly to account for the multiple CPIs at different frequencies, resulting in a test resembling the noncoherent sum of terms of the form in (10.99) for each of the frequencies, where the maximum is the estimate of the target angle and range-rate.

Information on the target's position and velocity is then generally fed into an automatic tracker.

10.7 | OTHER CONSIDERATIONS

This section briefly highlights several important issues in STAP implementation, each of which concerns the interference-plus-noise covariance matrix: the fidelity of the estimate in the presence of real-world effects; and computational considerations.

10.7.1 Heterogeneous Clutter

Equation (10.71) provides an estimate of the interference-plus-noise covariance matrix [15]. This approach is maximum likelihood if the training data, $\{\mathbf{x}_m\}_{m=1:P}$, are IID and multivariate Gaussian. The IID assumption means that training data come from the same distribution as the null-hypothesis condition of the cell under test and are independent samples. This is nominally the same as saying the training data are homogeneous. When the training data are homogeneous

$$E[\hat{\mathbf{R}}_k] = E\left[\frac{1}{P} \sum_{m=1}^P \mathbf{x}_m \mathbf{x}_m^H\right] = \frac{1}{P} \sum_{m=1}^P \mathbf{R}_m = \mathbf{R}_k \quad (10.100)$$

where $\mathbf{R}_m = \mathbf{R}_k \forall m$. In other words, the covariance matrix estimate converges to the true but unknown covariance matrix, \mathbf{R}_k , indicating the adaptive filter likewise converges to the correct response.

Real-world environments tend to appear heterogeneous. Heterogeneous clutter results from culturally varying features of the terrain and leads to space-time data violating the IID assumption. Examples of heterogeneous clutter include spatially varying clutter reflectivity such as at a land–sea interface; clutter discretized due to man-made objects; extended discretized including fence lines or buildings; and other target-like objects such as ground traffic. As a result of heterogeneous clutter, the covariance matrix estimate tends to the average characteristics of the data in the training interval, as (10.100) suggests, which may appear mismatched to any particular range bin of interest. As the covariance matrix estimate directly influences the adaptive filter frequency response, heterogeneous clutter can lead to mismatches in filter null depth, width, and placement.

Increased clutter residue and signal cancellation are two common impacts of heterogeneous clutter, resulting in increased false alarm rate and degraded detection probability. The impact of heterogeneous clutter on STAP performance and ameliorating solutions are discussed in detail in [26–28].

10.7.2 Nonstationary Angle-Doppler Region of Support

While it is a matter of semantics, the STAP community sometimes distinguishes between heterogeneous and nonstationary clutter conditions. Nonstationary clutter exhibits range variation in its angle-Doppler properties. As the clutter covariance estimate tends to the average properties of the space-time data comprising the training interval, as (10.100) suggests, nonstationary clutter conditions lead to incorrect null placement. Nonstationary clutter results from sensor collection geometry and is observed in forward-looking arrays operating at near range [4], in conformal arrays in other than sidelooking configurations [30], and in airborne bistatic radar topologies [31]. The STAP research community has devised a number of strategies to mitigate the impact of nonstationary clutter, including localized training, time-varying weightings, and data warping [30,31]. Of these techniques, data warping is the most popular and involves modulating the space-time data on a range bin basis to a reference point in an attempt at homogenizing the clutter covariance matrix properties over range; STAP is then applied directly to the compensated data.

It is possible that heterogeneous and nonstationary clutter coexist, further exacerbating the challenges associated with each.

10.7.3 Computational Burden

STAP remains a computationally demanding signal processing technique. Adaptive weight calculation involves either implicitly or explicitly solving for the covariance matrix inverse, a cubic operation in the processor's DoFs. The RD-STAP methods described in Section 10.6 greatly alleviate the computational demands of joint domain STAP and provide a pathway to real-time implementation. Nevertheless, the computational aspects of STAP limit weight updates and the implementation of certain techniques devised to mitigate heterogeneous or nonstationary clutter. This generally means that trade-offs in algorithm design become necessary to support real-time capabilities. Examples of typical trades include training over several large training blocks to minimize the number of adaptive weight updates, minimizing post-detection logic requiring recalculations of range-Doppler maps, and attempts to separately mitigate clutter and radio frequency interference to minimize the number of spatial DoFs.

10.8 | FURTHER READING

STAP has been one of the most researched topics in radar technology over the past 15 to 20 years. For this reason, numerous extant conference and journal papers are easily found by interested readers.

This chapter introduces the reader to important STAP fundamentals for clutter suppression. While the interested reader should consider consulting all the provided references, [4,5] are recommended as starting points. These full-length references give in-depth detail on STAP specifics.

Moreover, [8] has proven popular and should be worthy of quick review. Also, [3,15] are classic papers that should be read early in one's study of STAP.

10.9 | SUMMARY

Aerospace radar systems must detect targets competing with strong clutter and jamming signals. For this reason, the radar system designer incorporates a mechanism to suppress such interference. A familiarity with the application of space-time degrees of freedom—shown herein to greatly enhance detection in interference-limited environments—is thus essential. This chapter discussed the important role of space-time adaptive processing in moving target indication radar.

The chapter began with descriptions of spatial and temporal sampling, along with beamforming and Doppler processing. Both beamforming and Doppler processing methods were shown to implement matched filtering, a procedure leading to maximal signal-to-noise ratio. Joint space-time signals and characterization approaches based on the two-dimensional power spectral density, eigenspectrum, and minimum variance distortionless response spectrum were then discussed.

Leveraging these discussions, the chapter then introduced the space-time properties of ground clutter returns as the composite of the space-time responses of many clutter patches summed over all azimuth at the range of interest. This discussion described the clutter ridge resulting from a deterministic angle-Doppler coupling, thereafter providing a numerical example showing the region of the angle-Doppler spectrum occupied by the clutter response.

Following a fundamental description of spatial, temporal, and space-time signals, a generic space-time processing architecture was then developed, and critical performance metrics such as probability of detection, probability of false alarm, SINR loss, and improvement factor were introduced. It was shown that maximizing SINR equivalently maximizes the probability of detection for a fixed probability of false alarm. Hence, the chapter then developed the space-time weight vector leading to maximal output SINR and characterized the STAP as a data-domain implementation of this optimal filter. The minimum variance beamformer and generalized sidelobe canceller space-time filters were also discussed as alternate STAP implementations, each also maximizing output SINR.

In actuality, STAP includes a whole class of linear filters operating on space-time observations. Section 6 introduced the reader to reduced-dimension STAP, including detailed considerations of the EFA and JDL post-Doppler STAP methods. A performance comparison of these two methods to optimal and joint domain STAP capabilities using the SINR loss metric was given. Additional topics discussed included pre-Doppler STAP and reduced-rank STAP, an integrated approach to angle and Doppler estimation, and an end-to-end processing scheme.

The chapter concluded with a brief discussion of additional, important STAP considerations centered on the challenges of estimating and manipulating the interference-plus-noise covariance matrix: heterogeneous clutter, nonstationary clutter, and computational burden. These are important topics that will continue to demand the interest of the radar research community.

10.10 | REFERENCES

- [1] Skolnik, M.I., *Introduction to Radar Systems*, 2d ed., McGraw Hill, New York, 1980.
- [2] Johnson, D.H. and Dudgeon, D.E., *Array Signal Processing: Concepts and Techniques*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [3] Brennan, L.E. and Reed, I.S., "Theory of Adaptive Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 9, no. 2, pp. 237–252, March 1973.
- [4] Klemm, R., *Space-Time Adaptive Processing: Principles and Applications*, IEE Radar, Sonar, Navigation and Avionics 9, IEE Press, 1998.
- [5] Ward, J., *Space-Time Adaptive Processing for Airborne Radar*, Lincoln Laboratory Technical Report, ESC-TR-94-109, December 1994.
- [6] Klemm, R., Ed., Special Issue on STAP, *IEE Electronics & Communications Engineering Journal*, February 1999.
- [7] Melvin, W.L., Ed., Special Section on STAP, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, April 2000.
- [8] Melvin, W.L., "A STAP overview," *IEEE Aerospace and Electronic Systems Magazine—Special Tutorials Issue*, vol. 19, no. 1, pp. 19–35, January 2004.
- [9] Haykin, S., *Adaptive Filter Theory*, 3d ed., Prentice-Hall, Upper Saddle River, NJ, 1996.
- [10] Lathi, B.P., *Modern Digital and Analog Communication Systems*, 2d ed., Holt, Rhinehart and Winston, Inc., Philadelphia, 1989.
- [11] Marple, S.L., *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

- [12] Fenner, D.K. and Hoover, W.F., "Test Results of a Space-Time Adaptive Processing System for Airborne Early Warning Radar," in *Proceedings of the IEEE National Radar Conference*, Ann Arbor, MI, pp. 88–93, May 13–16, 1996.
- [13] Melvin, W.L. and Showman, G.A., "A Knowledge-Aided STAP Architecture," in *Proceedings of the DARPA/AFRL KASSPER Workshop*, Las Vegas, NV, April 14–16, 2003.
- [14] DiFranco, J.V. and Rubin, W.L., *Radar Detection*, Artech House, Dedham, MA, 1980.
- [15] Reed, I.S., Mallett, J.D., and Brennan, L.E., "Rapid Convergence Rate in Adaptive Arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 10, no. 6, pp. 853–863, November 1974.
- [16] Monzingo, R.A. and Miller, T.W., *Introduction to Adaptive Arrays*, John Wiley & Sons, New York, 1980.
- [17] Robey, F.C., Fuhrman, D.R., Kelly, E.J., and Nitzberg, R., "A CFAR Adaptive Matched Filter Detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, January 1992.
- [18] Chen, W.S. and Reed, I.S., "A New CFAR Detection Test for Radar," in *Digital Signal Processing*, vol. 1, Academic Press, pp. 198–214, 1991.
- [19] Griffiths, L.J. and Jim, C.W., "An Alternative Approach to Linearly Constrained Adaptive Beamforming," *IEEE Transactions on Antennas and Propagation*, vol. AP-30, no. 1, pp. 27–34, January 1982.
- [20] Goldstein, J.S., Reed, I.S., and Zulch, P.A., "Multistage Partially Adaptive STAP CFAR Detection Algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 2, pp. 645–661, April 1999.
- [21] Gabriel, W.F., "Using Spectral Estimation Techniques in Adaptive Processing Antenna Systems," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 291–300, March 1986.
- [22] Tufts, D.W., Kirsteins, I., and Kumaresan, R., "Data-Adaptive Detection of a Weak Signal," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, no. 2, pp. 313–316, March 1983.
- [23] DiPietro, R.C., "Extended Factored Space-Time Processing for Airborne Radar," in *Proceedings of the 26th Asilomar Conference*, Pacific Grove, CA, pp. 425–430, October 1992.
- [24] Wang, H. and Cai, L., "On Adaptive Spatial-Temporal Processing for Airborne Surveillance Radar Systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 660–670, July 1994.
- [25] Blum, R., Melvin, W., and Wicks, M., "An Analysis of Adaptive DPCA," in *Proceedings of the IEEE National Radar Conference*, Ann Arbor, MI, pp. 303–308, May 13–16, 1996.
- [26] Melvin, W.L., "Space-Time Adaptive Radar Performance in Heterogeneous Clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 621–633, April 2000.
- [27] Melvin, W.L., Guerri, J.R., Callahan, M.J., and Wicks, M.C., "Design of Adaptive Detection Algorithms for Surveillance Radar," in *Proceedings of the IEEE International Radar Conference*, Alexandria, VA, pp. 608–613, May 7–12, 2000.
- [28] Melvin, W.L., "STAP in Heterogeneous Clutter Environments," in *The Applications of Space-Time Processing*, R. Klemm, Ed., IEE Radar, Sonar, Navigation and Avionics 9, IEE Press, 2004.
- [29] Davis, R.C., Brennan, L.E., and Reed, I.S., "Angle Estimation with Adaptive Arrays in External Noise Fields," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-12, no. 2, pp. 179–186, March 1976.

- [30] Hersey, R.K., Melvin, W.L., McClellan, J.H., and Culpepper, E., "Adaptive Ground Clutter Suppression for Conformal Array Radar Systems," *IET Journal on Radar, Sonar, and Navigation*, vol. 3, no. 4, pp. 357–372, 2009.
- [31] Melvin, W.L., "Adaptive Moving Target Indication," Ch. 11 in *Advances in Bistatic Radar*, N. Willis and H. Griffiths, Ed., SciTech Publishing, Raleigh, NC, 2007.

10.11 | PROBLEMS

The interested reader should accomplish the following problem set in the order given using a suitable numerical simulation capability, such as the MATLAB programming environment.

1. Based on the discussion in Section 10.2, generate a matrix of space-time steering vectors covering $\pm 90^\circ$ in angle and ± 1000 Hz in Doppler, where each column is a space-time steering vector. The resulting matrix is known as the steering matrix. Assume an 11-element, uniform linear array with half-wavelength spacing receiving 32 pulses. Take the wavelength to be 0.3 m and the PRF to be 2 kHz. Choose an arbitrary space-time test steering vector (any column of the steering matrix, preferably starting with the steering vector at 0° angle and 0 Hz Doppler), manipulate into the form of (10.35), and take the two-dimensional FFT. Compare this result with the complex inner product of the steering matrix and the test vector. After evaluating several test cases, next duplicate Figure 10-5.
2. Using the steering matrix of problem 1, generate a simple, space-time covariance matrix comprised of receiver noise and three unity amplitude space-time signals with the following characteristics: (1) -200 Hz Doppler frequency and 20° DOA; (2) 200 Hz Doppler frequency and -20° DOA; and (3) 0 Hz Doppler frequency and 0° DOA. Set the receiver noise to 1 watt/channel. Using the steering matrix, calculate the MVDR spectra, thereby reproducing Figure 10-7.
3. Using the space-time covariance matrix from problem 2, generate 3520 IID data realizations by multiplying the matrix square root of the covariance matrix by unity power, complex white noise. If \mathbf{R}_{test} is the covariance matrix, the corresponding MATLAB command is $\mathbf{x} = \text{sqrtm}(\mathbf{R}_{\text{test}}) * (\text{randn}(N * M, 3520) + \text{sqrt}(-1) * \text{randn}(N * M, 3520)) / \text{sqrt}(2)$, where $N = 32$, $M = 11$, and 3520 IID realizations represents training support equal to 10 times the processor's DoFs. Using the IID data, calculate the periodogram estimate via (10.40).
4. Using the covariance matrix from problem 2 and the IID data from problem 3, calculate the eigenspectra for known and unknown covariance matrices of varying sample support. Regenerate the space-time covariance matrix and IID data with variable signal amplitude (to vary the SNR) and recompute the eigenspectra. Compare the eigenvalues to the calculated, matched filtered SNR. The eigenvalues should be equal to the integrated signal power plus the single-element noise power; alternately, the difference between a given eigenvalue and the noise floor eigenvalue level is the integrated SNR.
5. Take the eigenvectors from problem 4 corresponding to the dominant subspace (those eigenvalues above the noise floor). Then, apply the two-dimensional Fourier transform to each of the three dominant eigenvectors; use the space-time steering vectors generated in problem 1 to accomplish this task. Compare the resulting spectrum with

the PSD in Figure 10-5 and the MVDR spectrum in Figure 10-7. The transformed eigenvector responses should show peaks aligning with the location of the signal peaks in Figure 10-5 and Figure 10-7.

6. Using results from the prior problems, calculate the optimal and adaptive weights at the following Doppler frequencies: -200 Hz, 200 Hz, and 0 Hz. In each case, steer the array spatial response to 10° . Plot the resulting spatial patterns (for each Doppler frequency) and determine if the filter nulls reside at 20° , -20° and 0° , respectively.
7. Redo problem 6 after varying the amount of training data used to calculate the sample covariance matrix. Compare the adaptive pattern using 2, 4, 6, and 10 times the processor's DoFs. Focus on the pattern null depth applied to the signal sources and the adaptive filter sidelobe pattern. Null depth does not change appreciably, but the spurious sidelobe level comes down significantly with increasing training support.
8. Implement the weight calculation on the right-hand side of (10.92) using the eigen-decomposition of the known covariance matrix calculated in problem 2. Ensure the result equals the weight vector given by (10.66).
9. Calculate the PCI weight vector in (10.93) using the covariance matrix from problem 2. Compare the resulting weight vector with the weight vectors calculated in problem 8. This can be done by comparing elements of the weight vector and plotting the corresponding filter patterns.
10. Using the steering matrix of problem 1, generate a simple, space-time covariance matrix comprised of receiver noise and two unity amplitude space-time signals with the following characteristics: (1) -200 Hz Doppler frequency and 20° DOA, and (2) 200 Hz Doppler frequency and -20° DOA. Set the receiver noise to 1 watt/channel. Following problem 3, generate 3520 IID snapshots. Insert a space-time signal at index 1 of 3520 (for convenience) with the following characteristics: unity amplitude gain on each element, 0 Hz Doppler frequency, and 0° DOA. Next, apply (10.99) to estimate the signal DOA and Doppler.

Space-Time Coding for Active Antenna Systems

François Le Chevalier

Chapter Outline

11.1	Introduction	499
11.2	Colored Space-Time Exploration.....	500
11.3	Interleaved Scanning (Slow-Time Space-Time Coding).....	515
11.4	Code Selection and Grating Lobes Effects	517
11.5	Wideband MTI [12], [4].....	520
11.6	Conclusion	524
11.7	Further Reading	525
11.8	References	525
11.9	Problems	526

11.1 INTRODUCTION

Long range surveillance is a very demanding application for radar designers, and now generally requires active antennas (using multiple transmitters) for improved power budget, increased availability, and beams agility and adaptability.

For ground and maritime systems, maintainability and availability requirements drive a trend towards distributed transmission, either across rows or columns of the antenna, or even as fully active antenna radars. The waveforms used for detection and tracking are generally relatively narrow-band (lower than 10 MHz), but target classification or identification requires higher resolution, either in a stepped frequency (successive coherent bursts, with carrier frequency varying from burst to burst) or in a wide instantaneous bandwidth mode.

For airborne systems, fixed targets are detected with high resolution synthetic aperture radar (SAR), requiring instantaneous bandwidths ranging from 500 MHz to 1 GHz, and moving air and ground targets are detected with MTI modes using space-time adaptive processing (STAP) for clutter rejection – which requires multiple channels on receive.

These very general considerations mean that modern radars generally implement multiple channels and relatively wideband (meaning larger than 10% relative bandwidth) transmit/receive radio-frequency front-ends. Taking that into account, one must then determine the best antenna architecture, waveform, and scanning strategy for a specific surveillance mission, taking into account the very limited time on target (or dwell-time) available for surveillance of wide areas at long range.

The purpose of this chapter is to outline the main possibilities and to show that the simultaneous requirement for wideband and multiple channels opens the way to new beamforming techniques and waveforms, where different signals are simultaneously transmitted in different directions, for jointly coding space and time, and coherently processed in parallel on receive. Such concepts, first proposed and demonstrated by S. Drabowitch and J. Dorey [1, 2], should now be considered as mature techniques to be implemented in operational systems.

Basically, the main advantages to be gained are a *better extraction of targets* – especially slow targets – from clutter, multi-path, and noise, and a *better identification of targets*, obtained through longer observation times and wider bandwidths.

Colored space-time exploration (Section 11.2), interleaved scanning (Section 11.3), taking into account the important issues of target coherence and diversity gain will then be considered in the following sections. Special consideration of grating lobes issues will then give orientations for code selection (Section 11.4), and wideband MTI will be examined (Section 11.5). The conclusion will summarize the benefits of an optimized space-time management on transmit and receive, for surveillance MTI modes.

11.2 | COLORED SPACE-TIME EXPLORATION

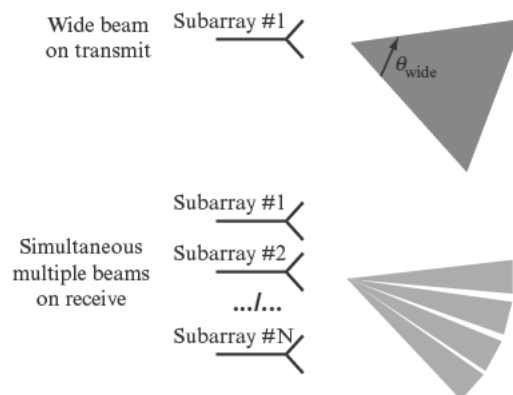
11.2.1 Digital Beamforming (DBF)

11.2.1.1 Principle and objectives

Standard digital beamforming is a procedure depicted in Figure 11-1, where a wide angular sector instantaneous coverage is obtained with a wide beam illumination on transmit (transmission through one sub-array), at least in one dimension – elevation or azimuth, and directive beams are formed on receive through coherent summations of signals received on different sub-arrays, in parallel for each direction.

Digital beamforming generally does not essentially change the power budget, compared to standard focused exploration, since the lower gain on transmit (due to wider illumination) is traded against a longer integration time (made possible by the simultaneous observation of different directions). Though precise performances depend on the fluctuation laws of targets and the requirements on detection and false alarms probabilities, one can take as a rule the global equivalence in power budget.

FIGURE 11-1 ■
Digital beamforming.



Digital beamforming may provide specific advantages, such as a better visibility of short events (e.g. RCS flashes), and a higher Doppler resolution especially useful for identification purposes, or for detection of slow targets, and effective suppression of noise jammers [4].

11.2.1.2 Limitations of Digital Beam Forming

However, for many applications, the wide illumination on transmit encounters different limitations, which may only be alleviated through simultaneous transmission of different signals in different directions – colored (space-time) transmissions.

For airborne applications, a severe limitation arises from the clutter spreading in Doppler,¹ due to the wider beam on transmit (which is, anyway, difficult to obtain with active antennas): this leads to a poor minimum detectable velocity, and to a poor clutter rejection, since only half the rejection is obtained (through the receiving antenna diagram), compared to focused beam illumination. This is one clear motivation for turning to colored transmission.

Another motivation, common to airborne and surface radars, comes from the fact that the waveform transmitted through the wide beam is of course the same for every direction in this beam. If, for instance, a surface radar uses a beam widened in elevation, in order to cover all altitudes simultaneously, the same waveforms will be transmitted for low elevation angles, where surface clutter is a primary limiting factor, implying the use of coherent bursts of pulses, and high elevation angles, where the target ranges are shorter and the surface clutter is largely reduced: since the optimal waveform depends on the elevation angle, it is clear that a widebeam with only one waveform is not optimal.

Moreover, when the mission is different for different elevation angles (e.g. detection of surface targets at 0° elevation, combined with detection of air targets at higher elevation angles), then again different waveforms are required in the different directions.

More generally, considering an active antenna with multiple receiving channels (i.e. multiple sub-arrays on receive), it can be seen that their use results in a difficult dilemma:

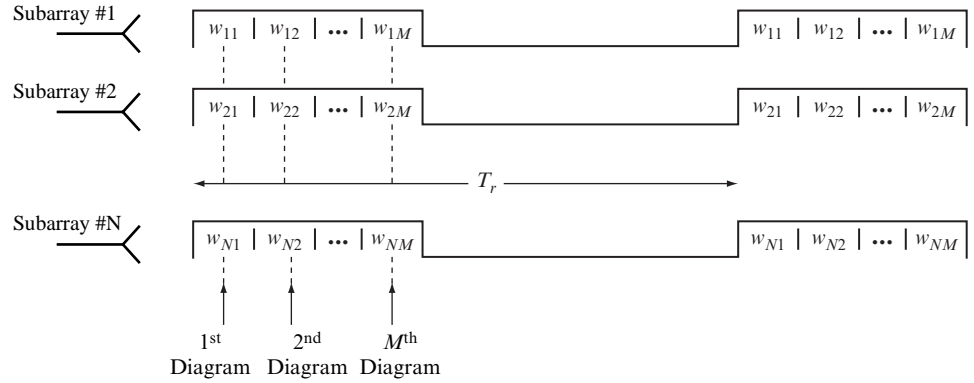
- If the beam is widened, then the above phenomena (clutter spreading in Doppler, or uniqueness of waveform) actually limit the overall performance;
- If the beam is focused, then the reception is not optimal, since each receiving channel has a wider sector of observation than the sector which is actually illuminated on transmit.

Space-time coding on transmit is a way to avoid this dilemma, by transmitting a wide beam while still providing angular directivity on transmit.

Last but not least, the directivity on transmit is also necessary when high angular resolution is required, for instance for mitigating multipath effects: the global directivity of the radar being the product of its directivities on transmit and on receive, a radar with directivity on transmit will have better performances against multipath than a wide beam system. Transmitting simultaneously different signals in different directions is the solution providing both wide angular coverage and optimal multipath rejection.

¹Main lobe clutter, as seen by a moving platform, has a spectrum whose main lobe is widened by the movement, especially for transverse observations (cf Vol 1, Chapter 5), since different scatterers on the ground have different radial velocities with respect to the platform. This widening of clutter spectrum is also observed when the platform is fixed, but the clutter is moving (atmospheric clutter, or sea clutter).

FIGURE 11-2 ■
Colored
transmission
(space-time coding).



11.2.2 Colored transmission, or space-time coding

11.2.2.1 Principles

The principle of colored transmission consists of simultaneously transmitting different waveforms in the different directions, thus achieving space-time coding² (Figure 11-2). In this figure, the coding is supposed to be a succession of M sub-pulses, coded in phase or frequency, but any type of code can be used [3]. For instance, transmitting different frequencies through the different sub-arrays could also be a possibility. The directivity on transmit is then recovered by signal processing on receive.

For signal processing on receive, the transmitted waveforms should be orthogonal, so that they can be separated from one another, on each receiving channel.

It should be emphasized that the transmitted waveforms are still periodic in time, since that is a necessary condition for an efficient cancellation of long range clutter (e.g. mountains: use of non-periodic waveforms is generally a very bad solution, since the far away returns from large structures, such as mountains or coasts, cannot be eliminated if they are not present in all the samples collected in the range gate – this is analyzed below, in Section 11.4).

Another way of considering such concepts is to describe them as the transmission, during each subpulse # m , through successive diagrams, the m^{th} diagram $D_m(\theta)$ resulting from the illumination law $w_{1m}, w_{2m}, \dots, w_{Nm}$ on the array, as illustrated in Figure 11-3 for three different concepts, which will be analyzed in more details later: frequency coding (identical diagrams at different carrier frequencies), fast angular scanning, and pseudo random orthogonal diagrams.

If the antenna is made of N sub-arrays arranged on a line Ox , $D_m(\theta)$ being the successive diagrams, and x_n the positions of the sub-arrays their analytical expression for a given coding sequence w_{nm} is written:

$$D_m(\theta) = \left| \frac{1}{N} \sum_{n=0}^{N-1} w_{nm} e^{-2i\pi \frac{x_n}{\lambda} \sin \theta} \right| \quad (11.1)$$

²This technique has been known as colored transmission, since the spatial distribution is then colored, as opposed to the white distribution corresponding to the wide beam. A more general denomination is space-time coding, meaning that the transmitted signals are now bi-dimensional functions, both of time and space.

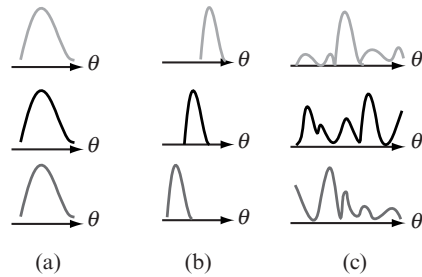


FIGURE 11-3 ■ Successive diagrams (a) Frequency coding (b) Fast angular scanning (c) Pseudo random orthogonal diagrams.

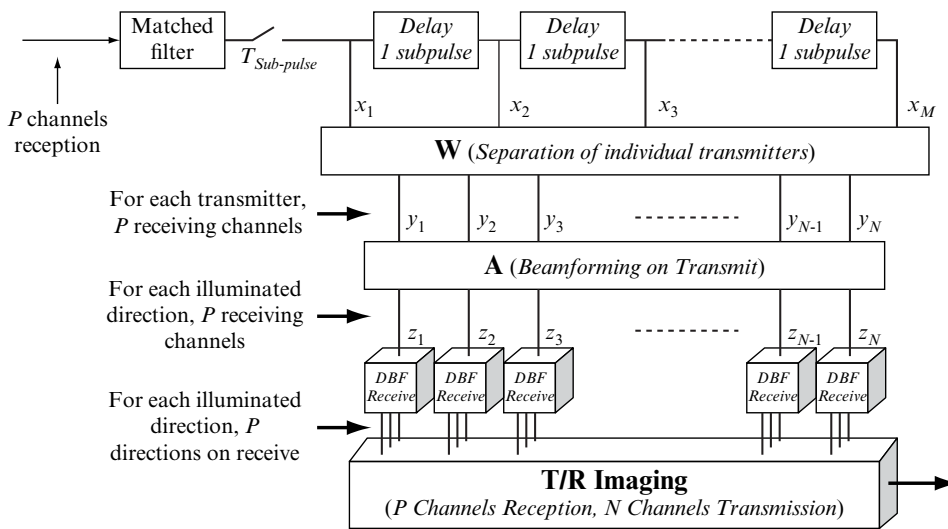


FIGURE 11-4 ■ Optimum reception of colored signals.

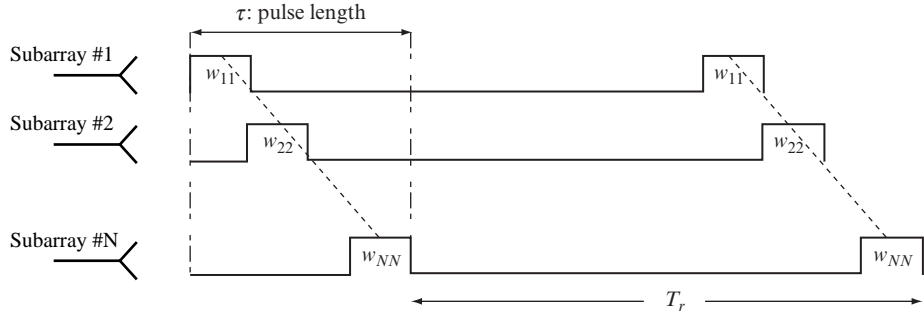
The optimum processing then basically consists in the operations described in Figure 11-4 (essentially a coherent summation of the received samples, for each angle – Doppler – range hypothesis):

- Transverse filtering, for separating the signals received from the different transmitters; For instance, if the transmitted signals are different frequencies, this is just a bank of filters separating the different frequencies; the y_n signals correspond to each transmitter n , for all the P receiving channels;
- Digital Beamforming on transmit (basically Fourier transform), coherently summing the transmitted signals, for each receiving antenna (i.e. each receiving channel); the z_n signals correspond to each direction n , for all the P receiving channels;
- Digital Beamforming on receive, for each direction on transmit (again basically Fourier transform);

Basically, there are then two beamforming algorithms which are used,³ one on transmit and one on receive – but not all the directions have to be examined, since for each transmitting direction only one receiving direction should be examined, in general.

³Excepted for intra-pulse scanning, as shown in the example below

FIGURE 11-5 ■
Circulating pulse.



Formally, if the same sub-arrays are used on transmit and receive ($P = N$), the expression to be evaluated is:

$$F(\sin \theta, \sin \theta_0) = \left| \frac{\sum_{k=0}^{N-1} \sum_{n=0}^{N-1} e^{2i\pi \left(\frac{x_k}{\lambda} \sin \theta - \frac{x_n}{\lambda} \sin \theta_0 \right)} \sum_{m=0}^{M-1} w_{km} w_{nm}^*}{\sum_{n=0}^{N-1} \left| \sum_{m=0}^{M-1} w_{nm} e^{2i\pi \left(\frac{x_n}{\lambda} \sin \theta \right)} \right|^2} \right| \quad (11.2)$$

For increased performance in cluttered environments or adverse conditions, digital beam-forming will preferably be performed with appropriate adaptive algorithms [4], on transmit and receive.

11.2.2.2 Circulating pulse

A simple example is illustrated in Figure 11-5: the “circulating pulse”, where a sub-pulse is successively transmitted through each sub-array. If the sub-arrays are regularly spaced (Uniform Linear Array) horizontally, this is equivalent to moving the phase center very rapidly through the whole array, thus creating an artificial Doppler (SAR effect) on transmit. For example, if the sub-pulse is 100 ns long, with 10 sub-arrays, this produces an artificial Doppler of ± 5 MHz (clearly distinct from the standard Doppler effect, which can only be measured as a phase shift from pulse to pulse). Formally, the code w_{kl} is written:

$$w_{nm} = \delta(n - m)$$

Note that in this example, $M = N$, i.e. the number of moments in the code (number of sub-pulses) is equal to the number of sub-arrays: this constraint need not be satisfied for all types of space-time codes.

The global effect is equivalent to a frequency coding in azimuth, which is shown in Figure 11-6, where each column represents the spectrum of the transmitted signal, evaluated through a Fourier transform on a duration equal to $1 \mu\text{s}$, duration of the global pulse (providing approximately 1 MHz resolution).

Referring to Figure 11-3 (space-time coding as a succession of diagrams), this coding is similar to type a): the diagram is identical from sub-pulse to sub-pulse, but the phase center of the antenna is changed (rather than the frequency, as supposed in 11-2).

In this example, it can be seen that the standard radar ambiguity is now a range-angle-Doppler ambiguity, since the coding is indeed a space-time coding. This effect is illustrated In Figure 11-7 – and, more specifically, the coupling between range and angle - : whereas the matched filtering for the “correct” range gate (gate where the target is present) correctly sums all the returns from a given target, while in the adjacent range gate (separated

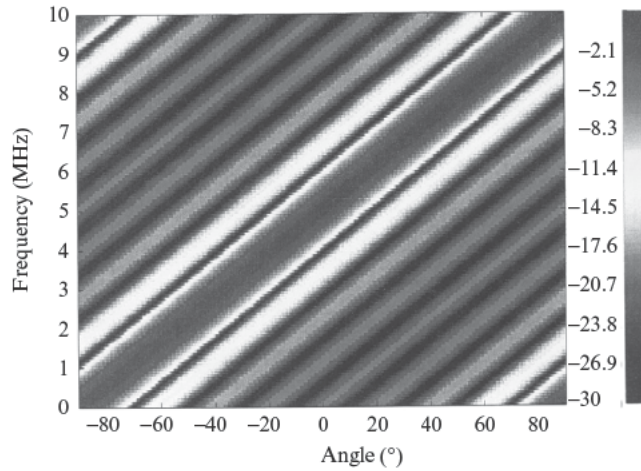


FIGURE 11-6 ■
Circulating pulse:
angle-frequency
coding sub-pulse:
100 ns long, with
10 sub-arrays on
transmit and only
one on receive.

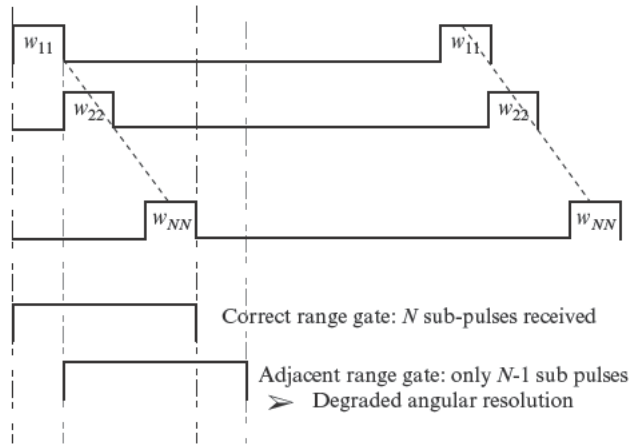


FIGURE 11-7 ■
Circulating pulse:
range-angle
ambiguity.

by $100 \mu\text{s}$ in our example), only $N - 1$ returns are summed, with a reduced effective antenna (by a factor $(N - 1)/N$) providing a degraded angular resolution by the same factor – and the same phenomenon with a factor $(N - 2)/N$ for the next range gate, and so on.

The resulting range-angle ambiguity function is shown in Figure 11-8, for the same example: the widening of the peak in angle, in adjacent ranges, is clearly visible.

The essential limitation of this simple space-time coding is the fact that only one transmitter is operated at each instant: it is generally preferable to use simultaneously all the transmitters, so as to maximize the effective radiated power (although that depends on the precise characteristics of the active elements, such as the maximum tolerable duty factor). The circulating code, presented below (Section 11.2.2.4), will alleviate this limitation.

Finally, it can be noted that this circulating pulse can also be used for “simulating” a continuous transmission with pulsed transmitters – a function that might be used for integrating pulsed radar systems in communication networks, for specific applications.

11.2.2.3 Fast scanning (intra-pulse scanning) [6]

In this mode, the angular diagram is rapidly scanned, from sub-pulse to sub-pulse, as described in Figure 11-3 (case b). There is a total ambiguity between time (range) and

FIGURE 11-8 ■
Circulating pulse:
range-angle
ambiguity (sub-
pulse: 100 ns long,
with 10 sub-arrays
on transmit and only
one on receive).

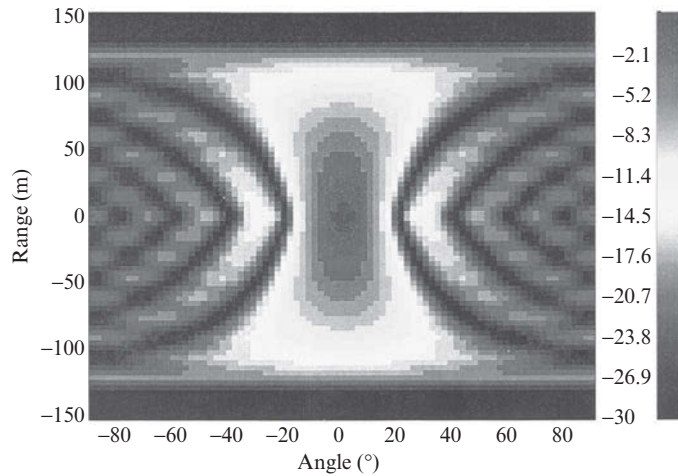
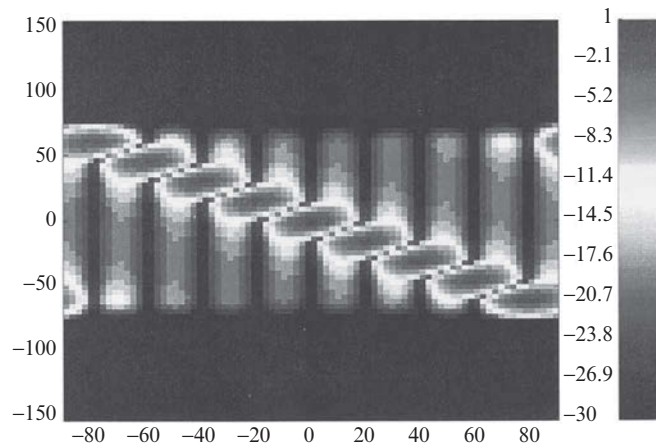


FIGURE 11-9 ■
Intra-pulse
scanning:
range-angle
ambiguity
(sub-pulse: 100 ns
long, with 10
sub-arrays on
transmit and only
one on receive).



angle, as shown in Figure 11-9, but this ambiguity can be removed, for example, by a symmetrical scanning in the opposite direction, or by changing the transmitted frequency from sub-pulse to sub-pulse⁴ – or, more simply, by Digital Beamforming on receive: this technique can be considered as a beam-widening technique on transmit.

An important advantage is that, compared to the other techniques, it *does not require any digital beamforming on transmit*, so the implementation is much less demanding in computer power. Variations of this technique are examined in Exercise 11.9.1

11.2.2.4 Circulating code [7],[5]

Another interesting example is given in Figure 11-10, where the coding is a “circulating chirp”: with 10 sub-arrays and 10 sub-pulses, 100 ns each, and one sub-array on receive: the operation is equivalent to a frequency-azimuth coding (one frequency in each direction).

⁴or, more generally, by coding the sub-pulses in phase or frequency (which means again a widened bandwidth, anyway)

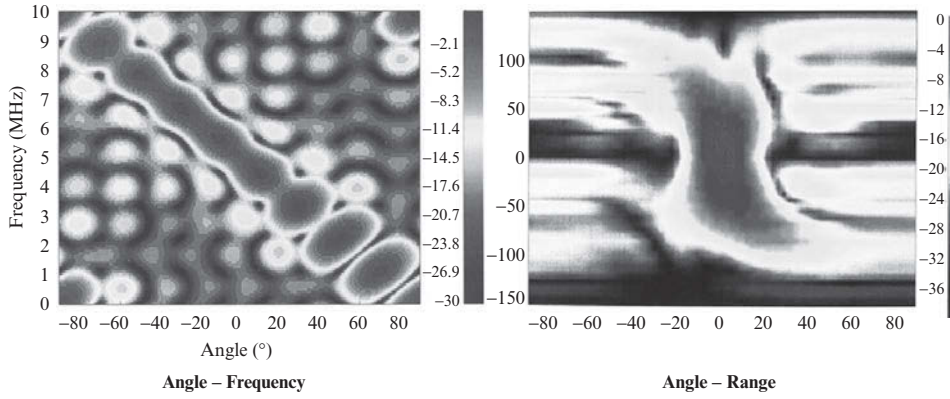


FIGURE 11-10 ■
Circulating chirp.

Formally, the code w_{nm} is written:

$$w_{nm} = e^{\frac{i\pi}{M}(n+m)^2} \quad (11.3)$$

Compared to standard DBF, such colored transmission schemes effectively provide angular separation on transmit, at the price of an increased instantaneous frequency bandwidth: higher Doppler resolution is now possible, without clutter spreading in Doppler.

Compared to the circulating pulse, the properties are similar, but all the available transmitters are now used simultaneously.

More generally, any type of circulating code, with the general property: $w_{nm} = w_{n+m}$, can be designed. Their basic property is that different frequencies are transmitted in the different directions. This property can be demonstrated, for a regular linear array, by observing that, for these codes, there is a time delay Δt between the signals transmitted by adjacent radiating elements; Consequently, since time delays translate as phase shifts in the frequency domain, this property imposes, at each frequency f , a phase shift $2\pi f \Delta t$ between adjacent sensors. This means that, at frequency f , the transmitted signals will be in phase for a direction of observation θ such that, if d is the distance between adjacent sensors:

$$2\pi f \Delta t = 2\pi \frac{d \sin \theta}{\lambda} \Rightarrow \sin \theta = \frac{\lambda}{d} f \Delta t \quad (11.4)$$

This means that there is a direct relation between the transmitted frequency and the transmission direction, as stated. The diagram at frequency f thus only depends on the array, not on the code itself – which can be a simple binary phase code, for instance: this opens the way to low cost electronic scanning systems, as emphasized in [5]. Exercise 11.9.3 illustrates some specificities of these circulating codes.

11.2.2.5 Bidimensional frequency coding [15]

Another example for fighter radars is illustrated in Figures 11-11 and 11-12, where a bidimensional angular coding is implemented.

The array is made of about 1000 elementary antennas (possibly grouped under sub-arrays arrangement), and the pattern of transmission shown in Figure 11-11 (3 columns transmitting frequency f_1 at elevation θ_1 , 3 adjacent columns transmitting frequency f_2 at elevation θ_2 , and again 3 adjacent columns transmitting frequency f_3 at elevation θ_3), is rapidly scanned horizontally through the array, thus realizing an azimuth coding through a circulating code technique similar to the ones described previously.

FIGURE 11-11 ■
Bidimensional
coding.

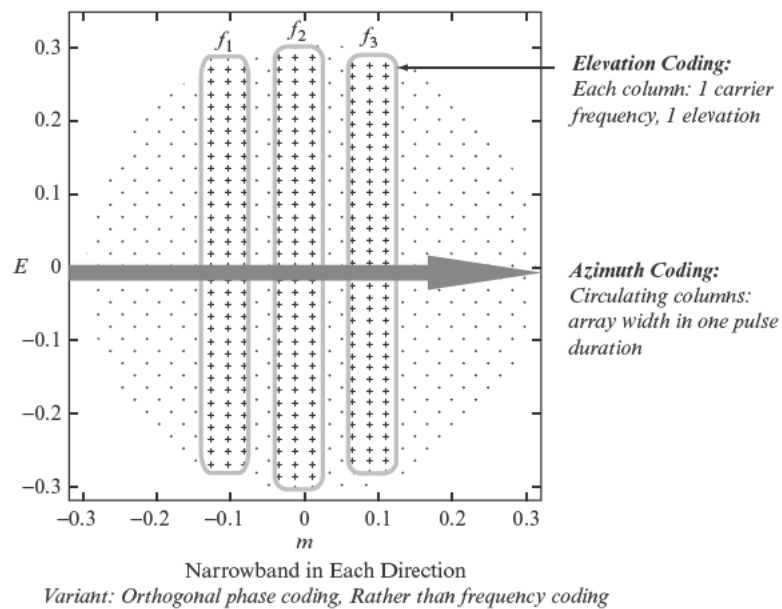
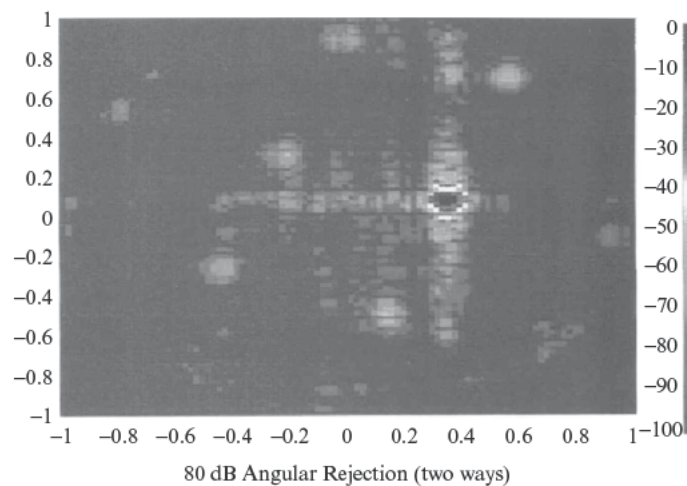


FIGURE 11-12 ■
Bidimensional
coding: 2 ways
diagram.



Globally, this is equivalent to an elevation/azimuth coding in frequency, providing on receive, after digital beamforming on transmit and receive,⁵ the diagram shown in Figure 11-12 for a specific angular steering direction: the sidelobe level is in the order of 80 dB (two ways), thus allowing Medium- or High- repetition frequency modes to be implemented for air-air detection and tracking over the global angular coverage. The overall extent of the angular sector can be adapted by suitable arrangements of the columns selected for transmission, taking into account the existing sub-arrays.

⁵Note that in the previous examples, only the transmitted diagram was analyzed (i.e. an omni-directional antenna was assumed on receive), whereas here, both transmission and reception beamforming has been performed.

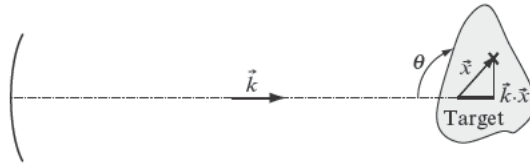


FIGURE 11-13 ■
Target scattering
coefficient.

11.2.3 Target coherence and diversity gains

Until now, the target was supposed to be an isotropic (in aspect angle) white (in frequency) scatterer, so that the received signals could be coherently added on reception. In reality, the target may more accurately be represented as a distribution of isotropic scatterers, characterized by their position \vec{x} relative to a specific point on the target, and by their complex diffraction coefficient $I(\vec{x})$.

This specific nature of the target has consequences on the performance of the radar system, since it changes the result of the coherent summation, and consequently the accuracy of the measurements, the fluctuation of the signals, etc. The following section will briefly summarize and illustrate the main results.

11.2.3.1 Target coherence

Within this multiple isotropic and white scatterers model,⁶ the scattering coefficient of the target, $H(\vec{k})$, can be written as a function of \vec{k} , the wave vector (vector along the incidence angle, modulus $2\pi/\lambda$), as described in Figure 11-13:

$$H(\vec{k}) = \int I(\vec{x}) e^{-2j\vec{k} \cdot \vec{x}} d\vec{x} \quad (11.5)$$

Since this expression is a Fourier transform, it can be inverted to provide the image $I(\vec{x})$ of the target (to within a scalar coefficient, obtained through calibration), based on the available measurements $H(\vec{k})$:

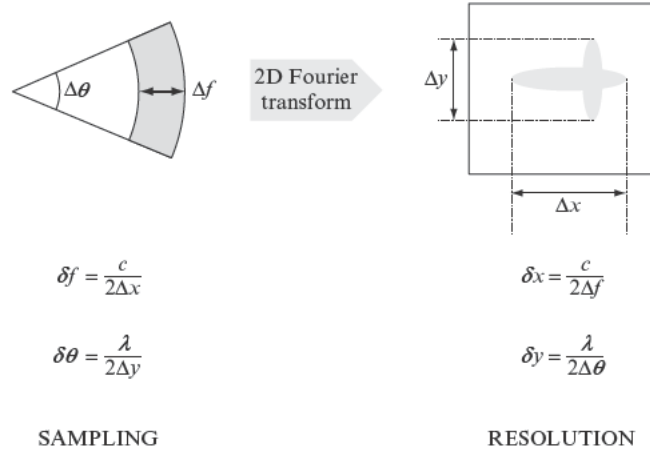
$$I(\vec{x}) = \int H(\vec{k}) e^{+2j\vec{k} \cdot \vec{x}} d\vec{k} \quad (11.6)$$

the limits of the integral being determined by the measurement system (usually a frequency bandwidth Δf around f_0 , and an angular sector $\Delta\theta$ around θ_0).

This is the basis of holographic measurements [4, 8], routinely used for target analysis. It also provides the basic parameters for sampling in the \vec{k} domain, and for resolution in the \vec{x} domain, when observing a target with depth Δx and transverse dimension Δy , with an observation bandwidth Δf associated with an angular sector $\Delta\theta$, as described in Figure 11-14 (these relations are mere consequences of the Fourier transform relationship between the measurements hologram $H(\vec{k})$ and the target image $I(\vec{x})$).

⁶In this case, “isotropic” means that the diffraction coefficient of the scatterer does not depend on aspect angle, and “white” means that it does not depend on frequency, *in the angular sector and bandwidth considered for coherent integration*. Though over-simplifying, this assumption is very generally used (e.g. for SAR and Inverse SAR imaging), and has been shown [4] to provide excellent quality images, even if some artifacts may be observed (resonances, or moving scatterers on the target, for example)

FIGURE 11-14 ■
sampling and
resolution criteria.



The image obtained on a real target drone (CT20) is presented in Figure 11-15a, for an observation in the horizontal plane, with an angular sector width of 20° around 40° and a frequency bandwidth of 2 GHz centered at 9 GHz.

Similar images, albeit with lower resolutions (typically metric), have been obtained on the field [20], as shown in Figure 11.15b for an airplane, and for a simulated ship (using the pitch angle variation)

These basic analysis tools provide essential parameters for target coherence: the critical instantaneous bandwidth and the critical antenna array extent.

The critical bandwidth Δf_c is defined as the maximum bandwidth such that the target coefficient remain coherent:

$$\Delta f_c = \frac{c}{2 \Delta x} \quad (11.7)$$

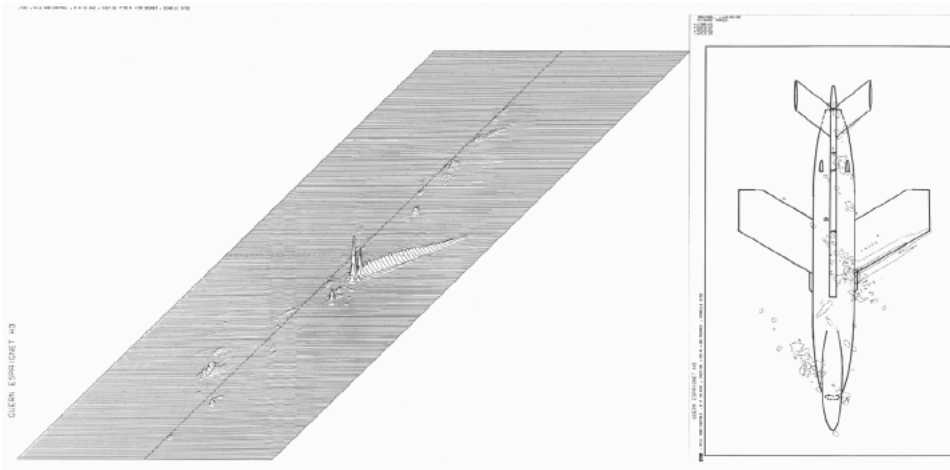
In other words, if the bandwidth transmitted *in the direction of the target* is equal to or larger than Δf_c , then the received signals can not be coherently summed – meaning that the target is in fact resolved in range by the signal. For instance, if the maximum dimension of the target is $\Delta x = 30$ m then $\Delta f_c = 5$ MHz.

As a consequence, when the bandwidth transmitted in the direction of the target is equal to or larger than Δf_c , the targets being resolved in range, some kind of distributed target integration has to be implemented, thus providing a diversity gain for fluctuating targets. This effect will be analyzed below (Section 11.2.3.2).

Similarly, as shown in Figure 11-16, the critical array extent D_c is the maximum extent of the antenna array such that the target coefficient, observed at range R , remains coherent:

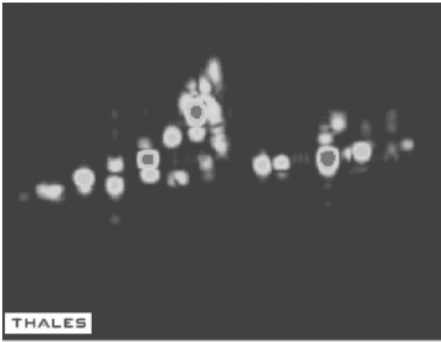
$$D_c = \delta \theta \cdot R = \frac{\lambda}{2\Delta y} R \quad (11.8)$$

If the extent of the array is larger than D_c , then the system cannot be considered as a monostatic system, and the received signals can not be coherently summed – in other words, the target is resolved in angle by the array (which again can provide some kind of diversity gain when correctly processed, see [18]). For instance, if $R = 100$ km, $\Delta y = 30$ m,



(a)

FIGURE 11-15 ■
 (a) Example of a real target (drone CT 20) image (Courtesy of ONERA).
 (b) ISAR-2D Image of Airbus (C-Band radar), and
 (c) Simulated ISAR-2D Image of ship using pitch motion (S-band) (from [20]).



(b)



(c)

$\lambda = 3$ cm we obtain $D_c = 50$ m: this means that very generally, for monostatic systems, the extent of the antenna array is much smaller than the critical extent, and only bistatic systems can be used for resolving the target in angle.

These two critical parameters Δf_c and D_c should be considered as basic parameters in the design of so-called MIMO systems, as was also correctly analyzed by V. Chernyak [9] and Y. Peng et al. [10,11] (“full rank observation matrix” being equivalent to using an inter-element spacing larger than D_c).

Coming back to the previous examples, it may be noted that the circulating pulse and the circulating chirp transmit only narrowband signals in each direction (these techniques are basically angle-frequency codings), whereas in the intra-pulse scanning technique a wideband signal (sub-pulse) is transmitted in each direction. This is an important aspect to take into account for system design, depending on the priorities given to target analysis characteristics, and electromagnetic compatibility, for instance.



FIGURE 11-16 ■
 Critical array extent D_c .

11.2.3.2 Diversity gain

Detection of radar fluctuating targets is known to be limited by the presence of noise, and by the fact that the target may provide only very small signals, for certain presentation angles or frequency of illumination.

If a high probability of detection is required, *some* non-coherent integration is preferable, not to get trapped in a low RCS zone, especially for highly fluctuating targets (SW 1). By *some*, it is meant that coherent integration must first be used, in order to get a sufficient signal to noise ratio (at least larger than one, so that it is not degraded too much by the non-linear operation).

This explains the “Golden Rule” which is often used: First improve S/N (coherent integration), then mitigate the low RCS zones (frequency agility on a few steps). The price to pay for that non-coherent integration (and the associated diversity gain) is lower Doppler resolution, because of shorter coherent bursts.

Turning to wideband radars, we first observe that the energy in the squared modulus of the impulse response is the same as the energy in the squared modulus of the frequency response, by Parseval’s theorem. So, summing the energy of the impulse response along the length of the target is equivalent, from a detection point of view, to summing the energy of the corresponding frequency response. Basically, integration along the range profile of the target (for a pre-assumed length of the targets of interest, say 15 m for air targets, for instance), is thus a way to combine coherent integration (used to obtain the range profile, with its associated Doppler spectra in each range cell) with non-coherent integration. In other words, for wide band radars, coherent integration time – and the clutter separation that it provides – needs not be reduced to take benefit of diversity gain.

The next question is: how many resolution cells on the target – or what is the best resolution for detection? The answer is given in Figure 11-17, showing the required S/N ratio *per cell*, as a function of the number of cells on the target: $N = 6$, or 10. The diversity gain can be defined as the difference between the required S/N per sample, between the coherent summation case and the non-coherent summation case: It clearly appears that the diversity gain is maximum for $N = 6$, between 2.6 and 7 dB, depending on the required probability of detection, for this case of Swerling case 1 and 2. There is still a gain for higher resolutions, but much smaller.

Similar analyses with Swerling Case 3 and 4 show that the gain is lower – as expected, since the fluctuations of Swerling 3 targets are smaller than those of Swerling 1 – but still exists, at least for probabilities of detection larger than 0.8: gain between 2.6 dB and 1 dB. The diversity gain would then become a loss for very high resolution of Swerling 3 targets (1 dB loss for a target analyzed in 30 cells and a required $P_d = 0.8$, not shown on the figure).

In conclusion, this analysis has shown:

- Even if the target RCS per range cell is reduced compared to the low resolution (standard) radars, the summation of the energies of adjacent range cells provide a diversity effect which more than overcomes this RCS reduction;
- The more the target fluctuates, the higher the diversity gain;
- Higher requirements in P_d lead to higher diversity gain;
- The smallest targets should be divided into 5–10 range cells, not more – which means that metric resolutions are sufficient;

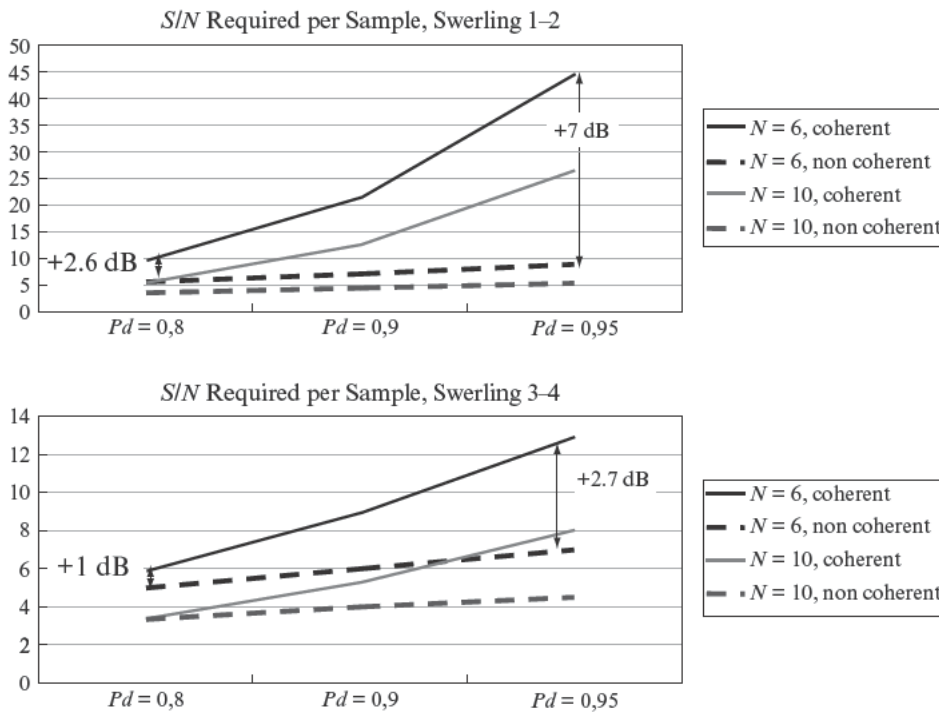


FIGURE 11-17 ■
Non-coherent versus
coherent integration
($P_{fa} = 10^{-6}$).

- For fluctuating targets, a few dBs can be gained from “cutting the targets into pieces”;
- *This diversity gain can be obtained without incurring the penalty of lower Doppler resolution, as would be the case with standard frequency agility.* In other words, with wideband radars, it is possible to keep the benefits of high Doppler resolution (e.g. improved slow target detection, target classification).

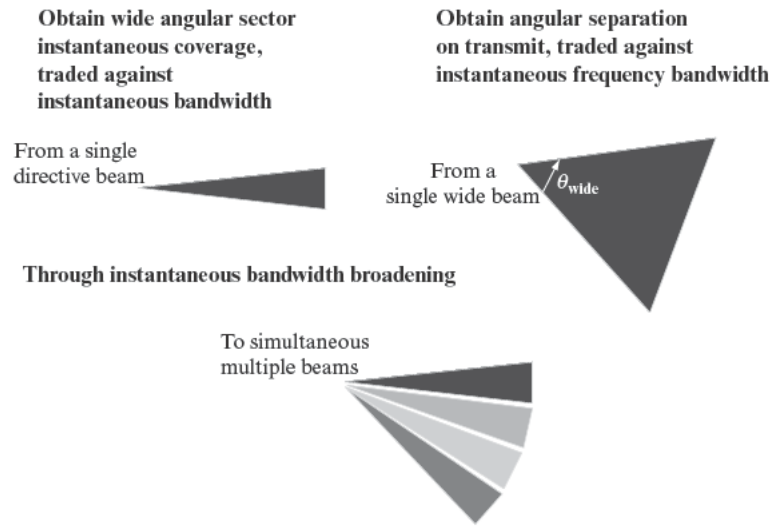
11.2.4 Colored transmission trade-offs and applications

As a whole, as shown in Figure 11-18, these colored transmission techniques can be described as providing angular instantaneous coverage (wide angular sector), traded against a larger instantaneous frequency bandwidth. Or they can be described, starting from the standard wide beam Digital Beamforming, as providing angular separation on transmit (hence better clutter rejection), traded against instantaneous frequency bandwidth.

A more provocative interpretation is the following: space-time coding is a way to improve the Doppler resolution (we have seen in 11.2.1.1 that this is the main benefit to be gained from widening the instantaneous coverage), at the price of an increased bandwidth, necessary to get multiple simultaneous transmissions, all other things staying equal. Improving the Doppler resolution by increasing the bandwidth certainly does not sound familiar to a standard radar expert: in this case, of course, increasing the bandwidth is just a way to increase the angular coverage – hence the Doppler resolution.

Such techniques could be used for instance for air-air combat mode, where they can provide the instantaneous wide coverage which is necessary while still maintaining a high visibility in clutter. Their strong resistance to jamming must also be emphasized, since any repeating jammer will give away its position by repeating the received code, thus

FIGURE 11-18 ■
Colored
transmission
trade-offs.



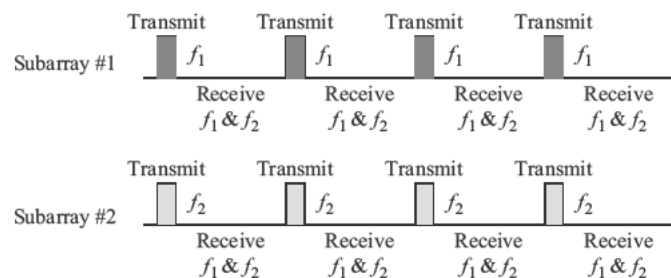
making it easy for the radar system to identify sidelobe jamming as such, and cancel the corresponding false plots.

These colored transmission techniques are also the solution to the standard “beams rendez-vous” problem for bistatic systems: allowing a wide beam on transmit without incurring the widening of the main beam clutter spectrum (since transmission directivity is recovered on receive), they provide the well-known benefits of bistatic systems (namely an improved detection in clutter through decreased clutter ambiguities, covertness, and ECCM).

For airborne surveillance radars, they also provide the solution to a classic dilemma: how to increase the Doppler resolution, needed for slow-targets detection and target classification, without widening the clutter spectrum. One possibility – actually very similar to the Radar à Impulsion et Antenne Synthétiques (RIAS) concept [2], [16] – is illustrated in Figure 11-19, where each sub-array transmits a different frequency carrier, coherently summed on receive as explained above. Depending on the specific system generation and reception constraints, it might be preferable to transmit different orthogonal codes (frequency modulations, phase codes, etc.), rather than different frequency carriers.

It must be emphasized that in this case, each target is illuminated by the whole bandwidth, thus making it possible to use such modes for High Range Resolution analysis of the detected targets, or wideband MTI, as described in Section 11.5.

FIGURE 11-19 ■
Colored
transmission for
surveillance.



11.3 | INTERLEAVED SCANNING (SLOW-TIME SPACE-TIME CODING)

Another way to explore space is obtained by interleaved scanning, where successive pulses are sent in successive directions, thus interleaving different pulse trains - possibly with different frequencies, or different codes. An example is shown Figure 11-20, with 2 interleaved directions.

This scheme, which can also be qualified as “slow-time” space-time coding, allows trading a wider quasi-instantaneous coverage – and the possibility to implement adaptive angular processing, by coherently processing the signals received from the adjacent beams – against a lower repetition frequency (and consequently more Doppler ambiguities) in each direction, and blind ranges or eclipses. It has no significant impact on the power budget: as for the previous space-time coding concepts, the loss in overall gain on transmit is balanced by a longer integration time on target.

With that interleaved scanning concept, it becomes possible to implement any adaptive procedure on receive, *with only one channel on receive*, if the transmitted signals are identical (so that the samples received from the different direction can be coherently processed to extract angle information). However, one has to take into account the fact that the samples are not taken simultaneously, so Doppler information has to be incorporated in the spatial filter. More specifically, the standard adaptive angular filter $W(\theta)$, to be applied to the vector z of collected samples in one range gate, which is classically written [4]:

$$y(\theta) = W^H(\theta)z \quad (11.9)$$

$$W(\theta) = \frac{R^{-1}a(\theta)}{\sqrt{a^H(\theta)R^{-1}a(\theta)}} \quad (11.10)$$

now becomes a Doppler-angle filter:

$$y(\theta) = W^H(\theta)z \quad (11.11)$$

$$W(\theta) = \frac{R^{-1}a(\theta, f_d)}{\sqrt{a^H(\theta, f_d)R^{-1}a(\theta, f_d)}} \quad (11.12)$$

$$a(\theta, f_d) = \Phi(f_d)s(\theta) \quad (11.13)$$

$$\text{with } \Phi(f_d) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & e^{2\pi j f_d T_r} & & 0 \\ \vdots & & \ddots & \\ 0 & & & e^{2\pi j (N-1) f_d T_r} \end{bmatrix} \quad (11.14)$$

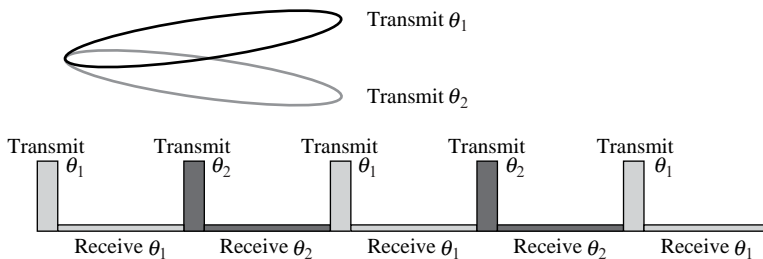


FIGURE 11-20 ■
Interleaved
scanning.

FIGURE 11-21 ■
Interleaved scanning
STAP.

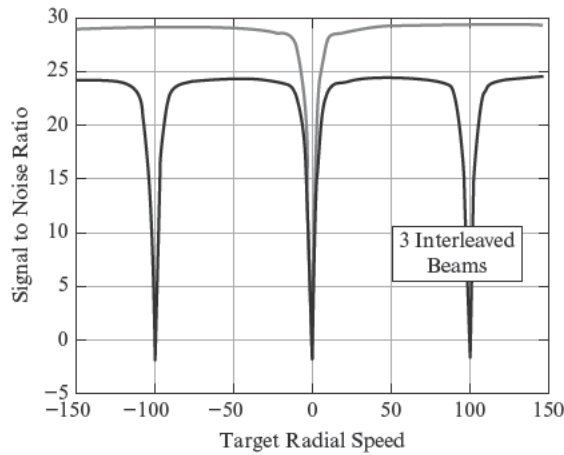
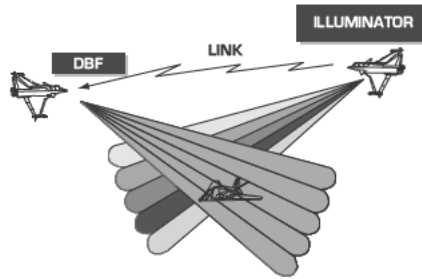


FIGURE 11-22 ■
Bistatic search and
track.

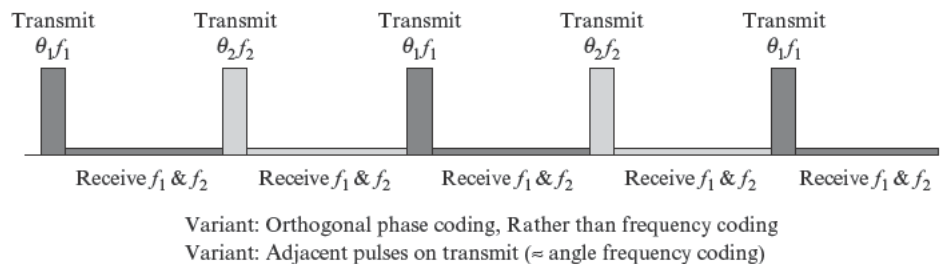


Such modes have been shown [12], [13], [14], to be an effective way of implementing space-time adaptive processing (STAP) with only one channel on receive: this is illustrated in Figure 11-21, with 3 interleaved directions. The result is of course an increased ambiguity in Doppler, and a loss of 5 dB in signal to noise ratio (due to the apparent widening of the beam), but a performance comparable to the 3 channels standard technique for slow targets detection.

This mode, as shown in Figure 11-22, is also a possible solution to the bistatic “beams rendez-vous” issue previously mentioned, providing the benefits of multistatic systems without increasing the completely of the synthesis/distribution parts, at the price of a reduced repetition frequency in each direction. Basically, that opens the way to multistatic search and track modes for low cost fighter radars, while effectively generating a medium repetition frequency in each direction.

Such modes could also provide efficient solutions for long range air-ground or air-air surveillance systems, preferably with adjacent pulses rather than evenly distributed – which amounts to trading the Doppler ambiguities against increased blind zone at short ranges (Figure 11-23).

FIGURE 11-23 ■
Interleaved scanning
Surveillance.



11.4 | CODE SELECTION AND GRATING LOBES EFFECTS

As is usually the case for radar, the design is primarily driven by clutter and jammer rejection constraints, either in the time domain – adequate rejection of clutter spectrum – or in the space domain – no grating lobes for better angular jammer (and moving clutter) rejection. Requirements for slow target detection and for target identification are also essential in modern air and surface radar design, and directly influence the strategy for space-time waveforms selection (maximizing the coherent integration time, the directivity on transmit, and the range resolution).

11.4.1 Time domain: periodic vs high time-bandwidth waveforms

In the time domain, there are basically two ways of rejecting clutter: use of a very wide time-bandwidth product, for obtaining a thumbtack range-Doppler ambiguity function with sufficiently low pedestal, or use of periodic pulses train, to obtain (through windowing) a sufficient Doppler rejection of main clutter lines. This is one powerful reason behind wideband signals, described in Section 11.5.

The necessity for periodic pulse trains is illustrated in Figure 11-24, where a sequence of irregularly spaced pulses is represented (greyed pulses). The black arrows, τ_t after transmission of the pulses, correspond to the samples received from a given target at range $(c/2)\tau_t$. If there is a *strong clutter echo at long range* (so-called “second trace” or second time around echoes), with delay τ_c after transmission of the pulse, as shown in red on the figure, it may happen that only one of these strong echoes falls in the range gate of the target – in the second pulse position, in this example. In this situation, it is not possible to suppress the clutter echo, since it is not present, in the same range gate, in response to the other pulses⁷.

It appears that only periodic pulse trains have the property that multiple strong second trace echoes can be properly cancelled by Doppler filtering. This is the main reason (beside the simplicity of generation and processing) behind the very general use of regular pulse trains for coherent radars.

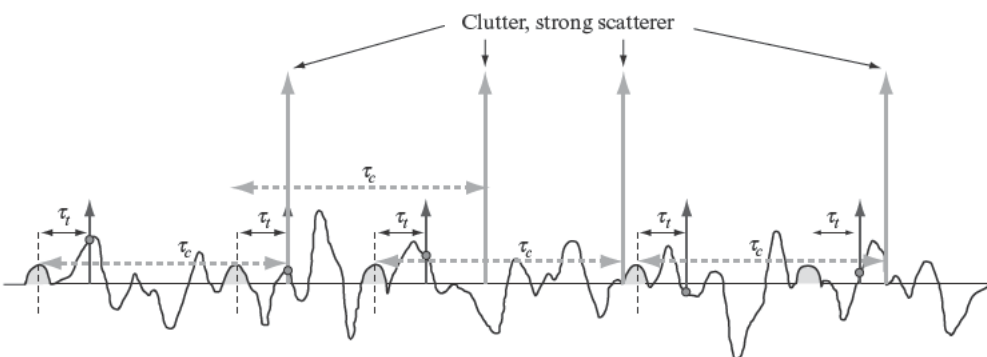


FIGURE 11-24 ■ Non periodic waveforms.

⁷Different tricks could be devised, for example suppressing the pulse whose amplitude is much larger than the other pulses in the range gate under consideration, or looking in other range gates to identify the clutter echo, etc. Generally, with strong echoes not completely reduced to point scatterers, the result will anyway be a degradation, compared to the periodic pulse train.

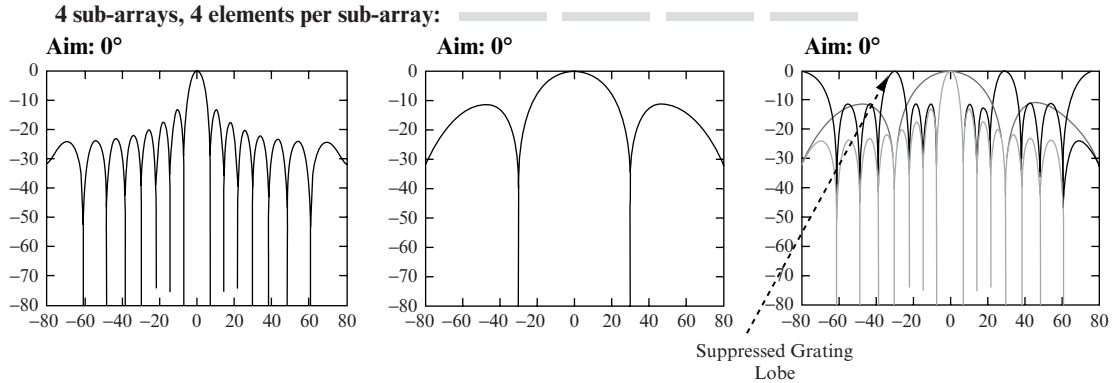


FIGURE 11-25 ■ Sub-arrays and grating lobes (left: ideal full array diagram; center: sub-array diagram; right: array factor with grating lobes, sub-array diagram, and final resulting diagram superposed) Sub-arrays and aiming oriented towards broadside direction.

11.4.2 Space domain: sub-arrays and grating lobes

In the space domain, the main difficulty is to minimize the number of channels – hence the number of sub-arrays – while still keeping the benefits of wide angle coverage. It will be seen that sub-arrays are a good solution for pencil beams observations, but are difficult to use for wide beam applications.

The phenomenon is best explained on an example, as shown in Figure 11-25, with a linear array of 4 sub-arrays, each constituted with 4 elementary radiating elements, 0.5λ apart from each other.

In Figure 11-25, are successively shown (from left to right):

- The diagram of the global array, each source being processed independently;
- The diagram of each sub-array of 4 radiating elements (with 0° phase shift on each radiating element);
- The resulting diagram (in red), when the array is obtained as the result of the 4 sub-array coherent summation, and when the phase shifts imposed on the 4 sub-array outputs correspond to an aiming direction equal 0° .

It can be seen that the resulting diagram, which is the product of the sub-array diagram (in green) by the diagram (in blue), of the under-sampled antenna made of the 4 phase centers of the 4 sub-arrays, is correct because the sub-array has a minimum exactly in the direction (30°) of the grating lobe of the under-sampled antenna (array factor constituted of the 4 phase centers of the 4 sub-arrays: blue curve).

However, looking in the direction 5° , for instance, as in Figure 11-26, then the phase shifts imposed on the outputs of the sub-arrays are changed: the result is that the grating lobes of the array factor (blue curve) do not correspond to the minimum of the sub-array diagram anymore (green curve), and there is a residual grating lobe, in red in Figure 11-26.

The way to mitigate this problem is of course to aim the sub-arrays in the aiming direction 5° , as shown in Figure 11-27: the grating lobe is now suppressed, as in the 0° aiming case.

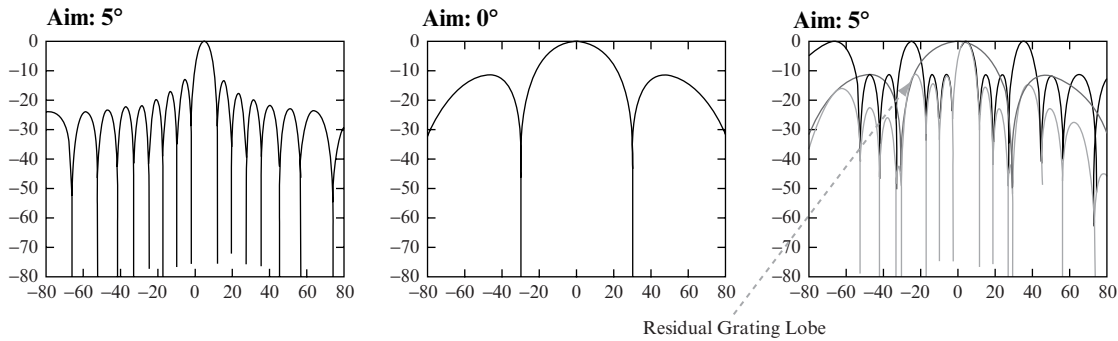


FIGURE 11-26 ■ Sub-arrays and grating lobes (left: ideal full array diagram; center: sub-array diagram; right: array factor with grating lobes, sub-array diagram, and final resulting diagram superposed) Sub-arrays oriented towards broadside direction, aiming in direction 5° .

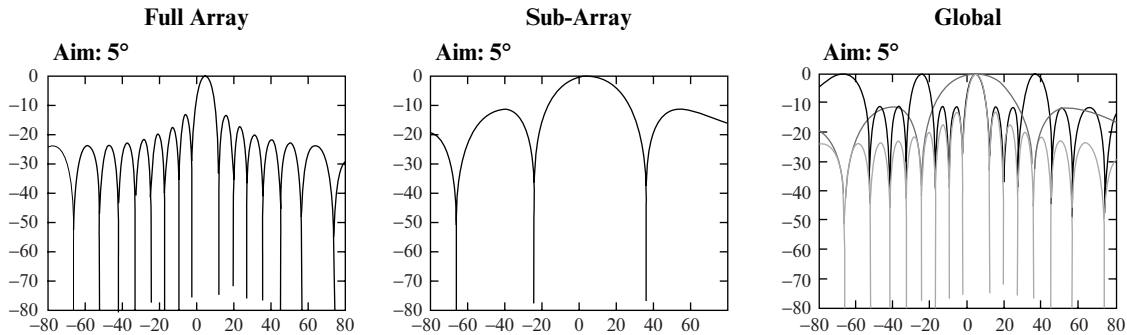


FIGURE 11-27 ■ Sub-arrays and grating lobes (left: ideal full array diagram; center: sub-array diagram; right: array factor with grating lobes, sub-array diagram, and final resulting diagram superposed) Sub-arrays oriented towards the exact aiming direction 5° .

The consequence for space-time coding is that, if the grating lobes cannot be tolerated,⁸ essentially three solutions remain:

1. Using interleaved scanning, where each transmission is made through a directive antenna – with corresponding aiming of the sub-arrays (we assume that there is a phase shifter behind each element in the array, as is the case in active antennas): this is a solution if the waveform is low repetition frequency, such that the signals are received, after each pulse, from the corresponding direction only – and then scanned for the next pulse;
2. Using intrapulse scanning, which also uses the transmit antenna only in a directive mode of operation (this is a solution only if the receiving antenna is correctly sampled);

⁸Depending on the application, such grating lobes may not be a real problem, if the receiving diagram is sufficiently clean. However, when the sub-arrays are identical on transmit and receive – which is the general case, but not the only one – the grating lobes are an essential limitation to wide angular coverage. Such wide beam modes are indeed more appropriate for one dimensional coverage with linear sub-arrays (adjacent columns or rows)

TABLE 11-1 ■ A large variety of space-time codings, which can be implemented either in fast-time (intrapulse) or in slow-time (from pulse to pulse)

		Code per transmitter		
		Frequency	Code	Time
Code per direction	<i>Frequency</i>	No	Circulating code	Circulating pulse (regular circulation)
	<i>Code</i>	RIAS: 1 frequency/transmitter, + random initial phase	Pseudo-random code	Circulating pulse (non regular circulation)
	<i>Time</i>	Regular array, 1 frequency/transmitter	Intrapulse scanning	No

3. Designing the sub-arrays in a non undersampled way, at least in one dimension – for instance per column or per row.

This simple analysis shows that the architecture of the array is not independent of the time waveform to be transmitted. Space-time coding involves the joint design of antenna illumination and time waveforms.

11.4.3 Coding strategy

The issue of space-time waveform selection is illustrated in Table 11.1, where different solutions are listed with their specific attributes. The codings are classified into three categories, by analogy with the communication field (Frequency Division, Code Division, and Time Division): frequency coding, phase coding, and time coding. Furthermore, each code can be conveniently described either in the space-time domain (a time code for each transmitter), or in the angle-time domain (a time code for each direction).

It must be emphasized that the most generic code (pseudo-random code) encompasses a wide variety of specific codes optimized for different desirable properties: low sidelobes in angle, or in range, or in a specific sub-domains of the range-angle domain.

This elementary representation of the wide diversity of space-time waveforms should also be combined with the antenna architecture (undersampled sub-arrays / correctly sampled sub-arrays / full active antenna) to be used as a guide for radar systems architecture design.

Furthermore, combinations of slow-time and fast-time codings could also be considered, for example fast-time coding in azimuth and slow-time coding in elevation, for shorter bursts at higher elevation angles.

11.5 | WIDEBAND MTI [12], [4]

An essential limitation for standard radars comes from pulsed radar range-Doppler ambiguity relation, which states that the ambiguous velocity V_a and the ambiguous range D_a are related by: $D_a \times V_a = \lambda \times c/4$. That relation means that many ambiguities, either in range or velocity (or both), have to be dealt with, which in turn implies the transmission of successive pulse trains with different repetition frequencies, requiring more time to be spent

on target for ambiguity and blind speeds removal (without corresponding gain in Doppler resolution, since the successive coherent pulse trains are then processed incoherently).

An alternative solution is obtained by increasing the range resolution, (or the instantaneous bandwidth) so that the moving target range variation (range walk, or range migration) during the pulse train becomes non-negligible compared with the range resolution – which is equivalent to stating that the Doppler effect is varying across the whole bandwidth (compared with the Doppler resolution), and can not be considered as a mere frequency shift any more – such radars may use bursts with low Pulse Repetition Frequency (no range ambiguities) wideband pulses such that the range walk phenomena during the whole burst is significant enough to remove the velocity ambiguity. It then becomes possible to detect the target and measure range and velocity with only one long coherent pulse burst.

The condition is written, if N_t is the number of pulses in the burst, T_r the repetition period, V_a the standard ambiguity velocity [$V_a = \lambda/(2T_r)$], ΔF the instantaneous bandwidth, and δR the range resolution [$\delta R = c/(2\Delta F)$]:

$$N_t V_a T_r \gg \delta R \Leftrightarrow \frac{\lambda}{2} N_t \gg \delta R \quad (11.15)$$

$$\Leftrightarrow N_t \gg \frac{F_0}{\Delta F} \quad (11.16)$$

For example, a burst of 60 pulses at 1 kHz repetition frequency with 500 MHz bandwidth in X band would be a possible candidate for non-ambiguous MTI detection. Similarly in S band, with 200 MHz, a burst of 45 pulses could be used with similar results.

The coherent signal processing of such radars (whose range resolution is in the order of a few wavelengths, typically less than 10) involves (Figure 11-28), for each velocity hypothesis, a coherent summation of the received echoes (Fourier transform), after range walk compensation:

$x_{r,t}$: received signal from p th pulse, at t th time sample

Hypothesis: range $t\delta R$, speed V

$$T_{t\delta R, V} = \sum_{p=0}^{N-1} x_{r, \Gamma} \left[t - p \frac{V T_r}{\delta R} \right] e^{-2\pi j p \frac{F_0}{F_r} \frac{2V}{c}} \quad (11.17)$$

with $\Gamma(u) =$ nearest integer from u

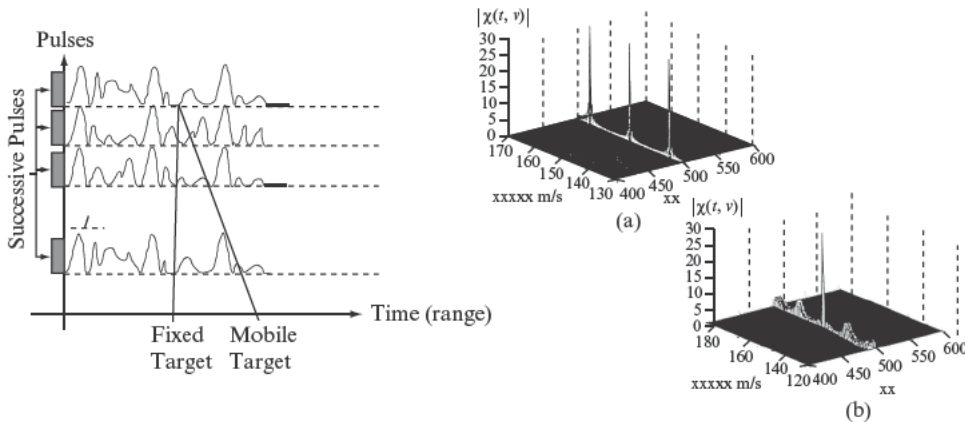


FIGURE 11-28 ■ Wideband signal processing, and Ambiguity functions (a: narrowband, 1/10000 bandwidth; b: wideband, 1/10 bandwidth).

More precisely, the matched filter for wideband MTI simply consists in a coherent summation of the received samples, for each possible velocity v and delay t ($t = 2R/c$) of the target. The literal expression is written [4]:

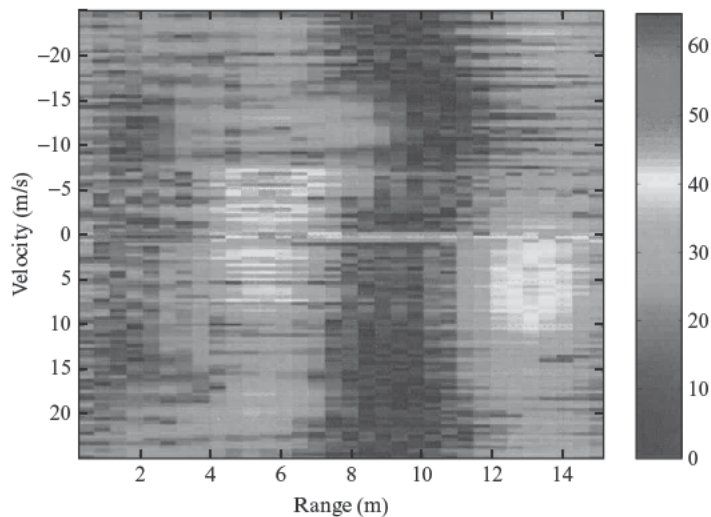
$$T(v, t) = \sum_{\substack{f=0, \dots, N-1 \\ r=0, \dots, M-1}} y_{f,r} \exp\left(2\pi j \frac{ft}{N}\right) \exp\left(-2\pi j r \frac{2v}{\lambda_0} T_r\right) \exp\left(-2\pi j r f \frac{\delta F}{F_0} \frac{2v}{\lambda_0} T_r\right) \quad (11.18)$$

In this expression, $y_{f,r}$ is the received signal as a function of frequency (sub-band) f and pulse number r , N is the number of frequencies (sub-bands), M is the number of pulses in the coherent burst, T_r is the repetition period, δF is the frequency step, F_0 is the central carrier frequency, and $\lambda_0 = c/F_0$ is the corresponding wavelength. The coupling between velocity and range, introduced by the migration effect (or, equivalently, by the fact that the Doppler shift is varying with frequency) is taken into account by the last term in this expression.

This processing leads to an ambiguity function (Figure 11-28) which does not exhibit the periodic ambiguities in Doppler, and provides the following advantages:

- Non-ambiguous detection of moving targets in clutter, with only one burst of high resolution pulses (possibly obtained with pulse compression, or stepped frequency techniques)
- Clutter attenuation (~ 20 dB, for a metric range resolution), due to the high range and Doppler resolutions;
- Simultaneous detection of fixed and moving targets (SAR + GMTI), with the high range resolution low PRF pulse train appropriate for SAR imaging ;
- High resolution range-Doppler Classification “on the fly”: Figure 11-29 illustrates this possibility with the image of a hovering helicopter, with 50 cm resolution, where the main rotor and the tail rotor are clearly visible at ranges 5 m and 13 m (signals obtained by Electromagnetic modelling of a Puma helicopter);

FIGURE 11-29 ■
Helicopter HRR
signature.



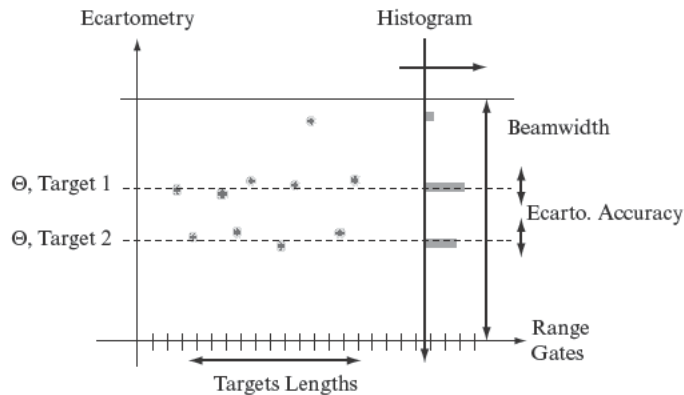


FIGURE 11-30 ■
Wideband
Ecartometry.

- ECCM properties (spread spectrum signals, requiring specific interception for ELINT or ESM, and specific devices for simulation of the wideband Doppler compression effect);
- Monopulse angular resolution of extended targets (essential for air-ground high target density situations), as shown in Figure 11-30: since an ecartometry measurement can be made in each range cell, the histogram of ecartometry measurements on an extended target will provide – if signal to noise ratio allows – the information that two targets are present, and their respective angular locations; Schematically, the angular resolution comes down to the – better – angular accuracy;

The performances of such wideband radars for moving targets extraction from clutter can even be improved, by subtracting “non-migrating” echoes, to keep only migrating targets: MiTI, Migrating Target Indicator [21], is a simple version of such technique, illustrated in Figure 11-31, for a radar at 10 GHz, $F_r = 1$ kHz, burst of 60 pulses, each with 250 MHz bandwidth. The principle consists in calculating a first image (Delay-Doppler), from the pulse train 1 to 57, then a second image, from the pulse train 3 to 60, and finally subtracting the amplitudes of these two images. This subtraction cancels stationary (non-migrating) targets, and migrating targets are then obtained as doublets (initial range – final range).

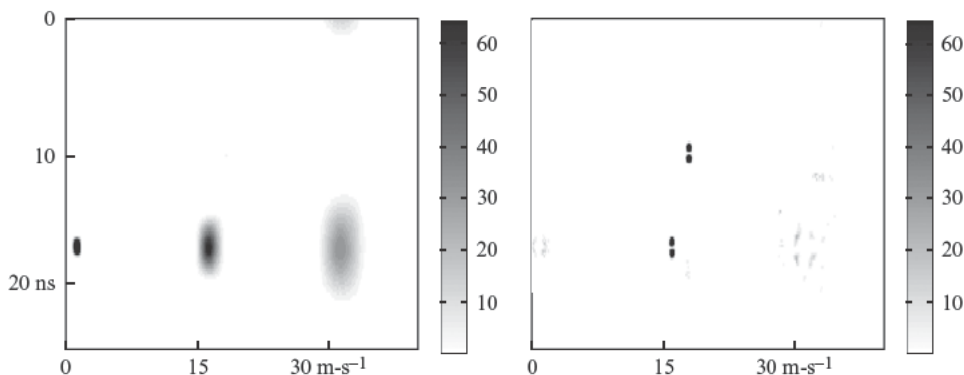


FIGURE 11-31 ■
MiTI, Migrating
Target Indicator Left:
standard
processing; Right:
Migrating Target
Indicator.

The literal expression of “Migration Target Indication”, which subtracts two images obtained by overlapping bursts: pulses 0 to $M - M_0 - 1$, and pulses M_0 to $M - 1$, is written:

$$T(v, t) = Abs \left\{ \sum_{\substack{f=0, \dots, N-1 \\ r=0, \dots, M-M_0-1}} [y_{f,r}] \exp \left(2\pi j \frac{ft}{N} \right) \exp \left(-2\pi j r \frac{2v}{\lambda_0} T_r \right) \exp \left(-2\pi j r f \frac{\delta F}{F_0} \frac{2v}{\lambda_0} T_r \right) \right\} \\ - Abs \left\{ \sum_{\substack{f=0, \dots, N-1 \\ r=M_0, \dots, M-1}} [y_{f,r}] \exp \left(2\pi j \frac{ft}{N} \right) \exp \left(-2\pi j r \frac{2v}{\lambda_0} T_r \right) \exp \left(-2\pi j r f \frac{\delta F}{F_0} \frac{2v}{\lambda_0} T_r \right) \right\} \quad (11.19)$$

In the example of Figure 11-27, the clutter is made of 2 slowly moving echoes (1 m/s and 0,5 m/s), with a delay of 17 ns (arbitrary origin), and two targets are inserted, 23 dB below clutter level, one at the same range as clutter (delay of 17 ns) et with velocity 16 m/s (slightly above the first ambiguous velocity of 15 m/s), the other exo-clutter at delay 10 ns and velocity 18 m/s.

On the left, only clutter echoes are visible, with residual ambiguous sidelobes around 16 m/s and 31 m/s, due to the repetition frequency 1 kHz (ambiguous velocity 15 m/s). On the right, these clutter residues are cancelled, leaving the two doublets corresponding to the two targets.

This simple example (which has then been validated with experimental trials on real clutter described in [22]) clearly illustrates the improvements that can still be obtained in clutter rejection with wideband radars – which naturally can be combined with space-time waveforms, to combine longer time on targets with wider bandwidths, and thus improve clutter and disturbances rejection.

11.6 | CONCLUSION

The different techniques briefly presented may be combined for simultaneous optimization of space exploration and target analysis and detection. For example, multifrequency transmission through different sub-arrays, or circulating chirp coding, with a total instantaneous bandwidth between 200 MHz and 500 MHz, or interleaved multifrequency pulse trains, will allow beamforming on transmit and receive and high resolution in range and Doppler, thus providing better detection, location, and classification of multiple surface and low-flying targets.

Basically, the main advantages to be gained are a *better extraction of targets* – especially slow targets – from clutter, multipath, and noises (including repeater jammers), and a *better identification of targets*, obtained through longer observation times and wider bandwidths.

More generally, it may be interesting to note that, while the second half of 20th century has seen major developments in radar waveform design and “Time/Doppler” signal processing, this early decade of 21st century is now focused on antenna developments around phased array design and “Space/Time” processing, on transmit and receive.

Cost reduction of Active Electronic Scanning Arrays and wideband integrated front-ends will enable generalization of these technologies for the more demanding applications, and intelligent radar management will of course be required to take full advantage of the bandwidth and agility available on surveillance radar systems.

11.7 | FURTHER READING

More details on optimization of space-time waveforms are given in the book *Waveform Design and Diversity for Advanced Radar Systems* edited by F. Gini, A. De Maio, L. K. Patton, Ed (Chapter 13: Space-time diversity for active antenna systems, by J.-P. Guyvarch, L. Savy and F. Le Chevalier). ISBN: 9781849192651; Publisher: IET, 2012

11.8 | REFERENCES

- [1] Drabowitch, S. and Aubry, C. "Pattern compression by space-time binary coding of an array antenna", in *Proc. AGARD CP 66*, Advanced Radar Systems, 1969.
- [2] Dorey, J., Blanchard, Y., Christophe, F., Garnier, G.: "Le projet RIAS, une approche nouvelle du radar de surveillance aérienne", *L'Onde Electrique*, vol 64, N° 4, 1978.
- [3] Levanon, N., and Mozeson, E.: *Radar Signals*, J. Wiley & Sons (Interscience Div.) New York, 2004.
- [4] Le Chevalier F., *Principles of Radar and Sonar Signal Processing*, Artech House, Boston, 2002.
- [5] Guyvarch J.P., "Antenne Spatio-Temporelle à Codes de Phases Circulants". in *Proc. Colloque GRETSI 97*, pp. 607–610, Grenoble, September 1997.
- [6] Calvary P. and Janer D., "Spatio-Temporal Coding for Radar Array Processing". in *Proc. ICASSP 98*, pp. 2509–2512, Seattle, 12–15 May 1998.
- [7] Le Chevalier, F.; Savy, L.; "Colored Transmission for Radar Active Antenna", in *Proc. International Conference on Radar Systems RADAR 2004*, Toulouse, France, October 2004.
- [8] Pouit C.: "Imagerie Radar à grande bande passante", in *Proc. International Colloquium on Radar*, Paris, 1978.
- [9] Chernyak, V.: "About the "New" Concept of Statistical MIMO Radar"; in *Proc. Third International Waveform Diversity & Design Conference*, Pisa, Italy, June 2007.
- [10] Dai, Xi-Zeng; Xu, Jia; Peng, Ying-Ning: "High Resolution Frequency MIMO Radar", in *Proc. CIE Radar 2006*, Shanghai.
- [11] Wu, Yong; Tang, Jun; Peng, Ying-Ning: "Analysis on Rank of Channel Matrix for Mono-static MIMO Radar System", in *Proc. CIE Radar 2006*, Shanghai.
- [12] Le Chevalier, F.; "Future concepts for electromagnetic detection", *IEEE Aerospace and Electronic Systems Magazine*, Volume 14, Issue 10, Oct 1999.
- [13] Le Chevalier, F.; "Smart beamforming and coloured signals for MIMO radars", Tutorial at the *Third International Waveform Diversity & Design Conference*, Pisa, Italy, June 2007.
- [14] Le Chevalier, F.; "Wideband ground surveillance waveforms", keynote speech at *AP SAR 2007* (in *Proc. CIE Asian-Pacific Synthetic Aperture Radar Conference*), Huangshan, PR China, November 2007.

- [15] Le Chevalier, F.; “Space-time transmission and coding for airborne radars”, *Radar Science and Technology*, vol 6, n° 6, dec 2008 (Bimonthly journal of CIE – Chinese Institute of Electronics).
- [16] Chen Baixiao; Zhang Shouhong; Wang Yajun; Wang Jun: “Analysis and experimental results on sparse-array synthetic impulse and aperture radar”, in *Proc. 2001 CIE International Conference on Radar*, Beijing, PRChina.
- [17] Le Chevalier, F.; Savy, L.: “Wideband coherent radar for moving target detection”, in *Proc. International Radar Symposium 98*, Munich, September 1998.
- [18] Stoica, P. and Li, J. Ed.: *MIMO Radar Signal Processing*, J. Wiley & Sons, New York, 2009.
- [19] De Maio, A., Lops, M. and Venturino, L.: “Diversity-Integration Trade-offs in MIMO Detection”, *IEEE Transactions on Signal Processing*, Volume: 56 Issue: 10, Oct. 2008,
- [20] Moruzzis, M., Ferrier, J.M., Gosselin, F., Enert, P., Lupinski, L. “MILORD: Synthesis and Perspectives”, in *Proc Radar 2004 International Conference*, Toulouse, October 2004.
- [21] Deudon F., Le Chevalier F., Bidon S., Besson O., and Savy L. “A Migrating Target Indicator for Wideband Radar”, in *Proc Sensor Array and Multichannel signal processing workshop, SAM 2010*, Israel, October 2010.
- [22] Le Chevalier F., Krasnov O., Deudon F., Bidon S. “Clutter suppression for moving target detection with wideband radar”, in *Proc 19th European Signal Processing Conference, EUSIPCO 2011*, Barcelone, September 2011.

11.9 | PROBLEMS

11.9.1 Fast scanning

Fast scanning, described in Paragraph 13.2.2.3, is a concept with many variations, either in fast-time or slow-time (interleaved) version, or both. In this exercise we will analyse some examples, for a scanning in elevation with a surveillance radar. The radar has 4 fixed panel antennas for 360° azimuth coverage, each antenna is made of 16 rows with independent transmitters and receivers, the typical repetition frequency is 1 kHz. The standard waveform uses pulses 10 μ s long, the beam is a fan beam, wide in elevation and narrow in azimuth – assume 100 beam positions in azimuth.

- (1) Regular fast-time scanning: Using the regular fast-time scanning, described in Paragraph 13.2.2.3, and with a global update rate of 3 s, evaluate the time on target, and the maximum number of pulses on target.
- (2) Optimisation of power budget: If maximum energy is not needed for all elevation angles, show that the optimisation of the sequence of fast scanning pulses can in effect result in an optimised distribution of energy in elevation, while keeping a constant duty factor for the transmitters.
- (3) Doppler filtering: Suppose that for high altitude targets only 3 pulses are required for clutter rejection, whereas for low-altitude targets more pulses are required. Assume that low-altitude waveform is required for the 2 lowest elevations (with 2 bursts for blind speeds removal). Which type of fast scanning would be appropriate? (*Hint: Use the maximum number of pulses on low elevation beams – Not only one solution to this problem*).

11.9.2 Equivalence: holography – SAR – Tomography

- (1) Show that holographic imaging, as described in Paragraph 13.2.3, is strictly equivalent to Synthetic Aperture Radar, when the trajectory of the radar is a circle centered on the target. (*Hint*: Consider stepped frequency SAR).
- (2) Interpret the sampling relations in \vec{k} domain (Figure 11-14) as non-ambiguity conditions in SAR image: δf giving the range ambiguity, and $\delta \theta$ the Doppler ambiguity, with Δx related to the range gate, and Δy related to the beamwidth. Give the formal expressions for these relationships.
- (3) Show that holographic imaging is a complex version of tomographic imaging. (*Hint*: use the “projection slice theorem” stating that the Fourier transform of the projection of an image $I(\vec{x})$ is a slice in the Fourier transform of the image, $H(\vec{k})$).

11.9.3 Circulating codes

As stated in Paragraph 13.2.2.4, if there is a delay Δt between the codes transmitted through adjacent elements of a uniform linear array (N elements, spacing $d = \lambda/2$ between adjacent elements), that translates to a differential phase shift between adjacent array elements at frequency f (in baseband):

$$\Delta\phi = 2\pi f \Delta t$$

This is a regular phase shift, from one radiating element to the next: the effect is to steer the array in the direction θ (relative to the normal to the array):

$$\begin{aligned} \Delta\phi &= 2\pi f \Delta t = 2\pi(d/\lambda) \sin \theta, \text{ with } \lambda = c/(f + f_0) \sim c/f_0 \\ f &= (d/\lambda)(1/\Delta t) \sin \theta, \text{ linear relation between } f \text{ and } \sin \theta \end{aligned}$$

- (1) Another interpretation: synthetic moving array. Show that this process can be interpreted as an array moving with a velocity $d/\Delta t$. What is the “synthetic Doppler effect” in direction θ ? What is the total synthetic Doppler bandwidth Δf ? What is the resolution required in frequency, to recover the directivity of the antenna array? Compare the number of independent directions, $\Delta f/\delta f$, with the number of independent samples in the code, $T_{tot}/\Delta t$.
- (2) Code selection. What are the requirements on the code, in terms of length and number of moments, in relation with the antenna length?
- (3) Diagram quality. The transmission diagram is obtained by Fourier transform of the signals received on the array, at each time sample (and then eventually digital beam forming on receive, if an array of receivers is available). Show that a one bit PSK code (e.g. Barker code, or constant amplitude zero autocorrelation code) may thus provide an acceptable angular diagram.

Electronic Protection

Aram Partizian

Chapter Outline

12.1	Introduction	529
12.2	Electronic Attack	533
12.3	EW-Related Formulas	545
12.4	EP Overview	553
12.5	Antenna-Based EP	554
12.6	Transmitter-Based EP	561
12.7	Exciter-Based EP	562
12.8	Receiver-Based EP	567
12.9	Signal Processor-Based EP	572
12.10	Data Processor-Based EP	576
12.11	Summary	581
12.12	Further Reading	584
12.13	References	584
12.14	Problems	585

12.1 | INTRODUCTION

Any radar used for a military application is a potential target of hostile electronic warfare (EW) by an adversary force. Electronic warfare refers to the broad range of military actions intended to degrade or exploit an adversary's use of the electromagnetic spectrum. These include electronic attack (EA), electronic support (ES), and electronic protection (EP) [1]. Electronic attack, generally referred to as jamming, includes active and passive responses that alter the return signals being received by the radar from the environment in an attempt to degrade or deny radar operation. Electronic support includes passive sensing used to detect, locate, characterize, and identify threat radar emissions for the purpose of intelligence gathering, real-time situation awareness, or EA response. Electronic attack and ES, therefore, are generally directed at an opponent's radar. Electronic protection constitutes the radar's defenses against hostile EA and ES actions. The categories of EA, ES, and EP are roughly equivalent to the previously used terms electronic countermeasures (ECM), electronic support measures (ESM), and electronic counter-countermeasures (ECCM), respectively, although the latter terms still remain in common use to this day [2–6]. The battle of the electromagnetic spectrum is a never-ending contest of action and reaction,

in which superior knowledge of an opponent's capabilities is paramount to the ultimate outcome.

The purpose of this chapter is to provide a brief introduction to the topic of EP with primary emphasis to its application to countering active EA. The field of EP is broad and complex and well beyond the scope of this chapter, even in an introductory form. Specific details associated with fielded EW capabilities are necessarily sensitive and therefore excluded from open publication. It is possible, however, to gain insight into the diversity of the topic through exposure to some of the fundamental concepts and examples presented in this chapter. Several books available in the public domain, listed at the end of the chapter, provide a very thorough treatment of the field of EW in general as well as EP in particular.

12.1.1 Organization

The subsequent sections of this chapter are organized as follows. Section 12.2 presents a brief overview of EA. Section 12.3 presents formulas for analyzing relative strength of jammer energy to target signal energy for different types of jammers. Section 12.4 provides a top-level overview of EP. Sections 12.5 through 12.10 present a sampling of common EP concepts, categorized based on their most closely associated radar subsystem—antenna, transmitter, exciter, receiver, signal processor, and data processor. Section 12.11 summarizes the various EP concepts described and lists examples of applicable EA techniques. Sections 12.12 and 12.13 provide recommendations for further reading and specific references cited in the text. Section 12.14 provides exercises that reinforce some of the key concepts presented in the chapter.

12.1.2 Key Points

- Electronic attack is intended to degrade or deny radar measurements in range, Doppler, and angle dimensions through noise and deceptive interference.
- Noise jammers reduce radar sensitivity, while coherent jammers, by preserving radar phase information, inject false targets and erroneous track information.
- An important measure of the jammer effectiveness is the ratio of jamming energy to target signal energy in the radar detection cell—the jam-to-signal ratio.
- Radar EP is potentially implemented anywhere in the radar system including the antenna, transmitter, exciter, receiver, signal processor, and data processor.
- Radar EP techniques generally attenuate or separate the jamming relative to the target signal in various signal domains, including spatial, temporal, and spectral.
- Common EP features include sidelobe attenuation, waveform diversity, high average radiated power, wide dynamic range, range resolution, and track consistency tests.
- The diversity and complexity of EW requires careful consideration of the anticipated EA and ES operational conditions in selecting the appropriate EP design for a radar.

12.1.3 Notation

Key variable names and symbols used in this chapter are defined in the following list.

Symbol	Definition
--------	------------

B	radar waveform bandwidth
B_j	jammer noise bandwidth

c	speed of light = 3×10^8 m/s
d_j	jammer duty factor
F	radar receive noise factor
f_c	radar carrier frequency
f_D	Doppler frequency
f_p	radar pulse repetition frequency (PRF)
$g_{j,el}$	jammer electronic gain between receive and transmit antenna ports
$G_{j,rx}$	jammer receive antenna gain at the radar transmit polarization
$G_{j,tx}$	jammer transmit antenna gain at the radar receive polarization
G_r	radar receive antenna gain
G_{rj}	radar receive antenna gain toward jammer
G_t	radar transmit antenna gain
G_{tj}	radar transmit antenna gain toward jammer
J	jammer power (or energy) in radar detection bandwidth
k	Boltzman's constant = 1.38×10^{-23} joules/K
L_s	radar system losses (defined ≥ 1)
N	thermal noise power
n	number of radar pulses coherently integrated
P_j	jammer peak transmit power
P_r	radar peak received power
P_t	radar peak transmit power
R	range from radar to target
R_j	range from radar to jammer
S	signal power (or energy) received from a target return
T	reference noise temperature (typically 290 K)
T_{cpi}	radar coherent processing interval (CPI) time
v	velocity (radial velocity or range rate)
Δ	monopulse difference pattern or received voltage
λ	radar wavelength
Σ	monopulse sum pattern or received voltage
σ	target radar cross section
τ	radar pulse width

12.1.4 Abbreviations

Key abbreviations used in this chapter are defined in the following list.

Abbrev.	Definition
ADC	analog-to-digital converter
AGC	automatic gain control
AM	amplitude modulation
ARM	antiradiation missile
BBN	blinking barrage noise
BN	barrage noise
BSN	blinking spot noise
CBN	cooperative blinking noise
CFAR	constant false alarm rate
CONSCAN	conical scan

CPI	coherent processing interval
CW	continuous wave
DAC	digital-to-analog converter
dB	decibel
dBi	dB relative to isotropic
dBm	dB relative to 1 milliwatt
dBmi	dBm from an isotropic antenna
dBW	dB relative to 1 watt
DN	doppler noise
DRFM	digital RF memory
EA	electronic attack
ECCM	electronic counter-countermeasures
ECM	electronic countermeasures
EMI	electromagnetic interference
EP	electronic protection
ERP	effective radiate power
ES	electronic support
ESA	electronically scanned array
ESJ	escort (support) jammer
ESM	electronic support measures
EW	electronic warfare
FM	frequency modulation
ft	feet
g	acceleration due to gravity (9.8 m/s^2)
GHz	gigahertz
HOJ	home on jam
HPA	high power amplifier
Hz	hertz
IBW	instantaneous bandwidth
IDR	instantaneous dynamic range
IF	intermediate frequency
IG	inverse gain
JNR	jam-to-noise ratio
JSR	jam-to-signal ratio
km	kilometer
LET	leading-edge track
LFM	linear FM
LO	local oscillator
m	meter
m/s	meters per second
MHz	megahertz
MLC	main lobe cancellation
ms	millisecond
MTI	moving target indication
NCP	noise cover pulse
ns	nanosecond
PM	phase modulation
PRF	pulse repetition frequency

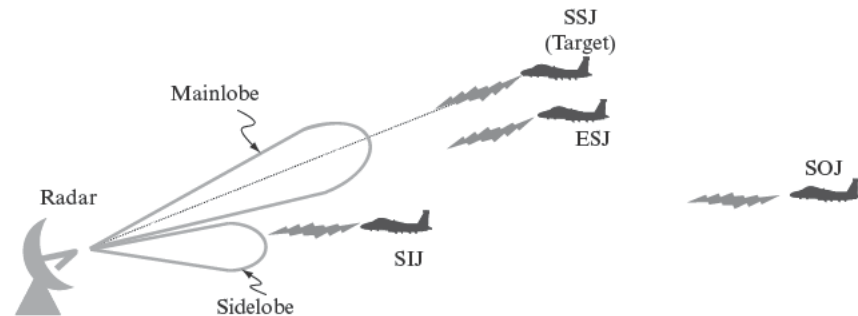
PRI	pulse repetition interval
PW	pulse width
RBM	range bin masking
RCS	radar cross section
RF	radio frequency
RFT	range false target
RGC	range gate capture
RGPO	range gate pull-off
RGS	range gate stealing
RSN	responsive spot noise
s	second
SAM	swept AM
SIJ	stand-in jammer
SINR	signal-to-interference-plus-noise ratio
SLB	sidelobe blanker
SLC	sidelobe cancellation
SLFT	sidelobe false target
SN	spot noise
SNR	signal-to-noise ratio
SOJ	stand-off jammer
SSJ	self-screening jammer
T/R	transmit/receive
TB	terrain bounce
TED	total energy detector
TOA	time of arrival
TRD	towed RF decoy
TWT	traveling-wave tube
VBM	velocity bin masking
VFT	velocity false target
VGPO	velocity gate pull-off
VGS	velocity gate stealing
W	watts
XEYE	cross-eye
XPOL	cross-polarization (cross-pol)
μ s	microsecond

12.2 | ELECTRONIC ATTACK

12.2.1 Electronic Attack Overview

Electronic attack is typically used to protect a target from being detected and tracked by opposition force radars. This is accomplished through two approaches: masking and deception. Masking is primarily intended to prevent detection or deny accurate measurement of target information such as range or Doppler through the generation of noise-like interference in the radar's detector. Deception is intended to inject false information into the radar processor through the generation of target-like signals that either create distraction false targets or surreptitiously draw the radar's range, Doppler, or angle track filters

FIGURE 12-1 ■
Radar–EW
engagement
geometries.



away from the true position of the target being protected. The line between masking and deception becomes blurred when the jammer overloads, or saturates, the radar with huge quantities of false targets, thereby delaying or preventing detection of the true target and effectively desensitizing the radar.

Electronic attack is implemented in either a support-jamming role, in which the jammer is deployed on a separate platform from the target being protected, or screened, or a self-protection jamming role, in which the jammer is located on the protected target platform itself [1,2,7]. Figure 12-1 illustrates the primary geometries of the support and self-protection jammer roles. A stand-off jammer (SOJ) is a support jammer that is at a much farther range from the victim radar than the target being screened, typically out of threat weapon range. The SOJ typically operates in the radar sidelobes as it is difficult to maintain angle alignment with the varying radar-target geometry and also because it may be screening multiple targets. The SOJ generally requires high power to overcome the large standoff range and sidelobe attenuation. A stand-in jammer (SIJ), or stand-forward jammer (SFJ), is a support jammer that is at a much closer range to the victim radar than the protected target. The SIJ generally requires less power than a SOJ due to the closer range, which both reduces the range-dependent loss and allows the jammer time to match its response to the current radar emission in a timely fashion. An escort support jammer (ESJ), or simply escort jammer, is a support jammer that flies in close proximity to the target being defended and is therefore at approximately the same range and angle as the target. Finally, a self-screening jammer (SSJ), or self-protection jammer (SPJ), has EA capability on the target platform itself. The EA equipment may be installed internally within the target platform or may be mounted as a pod on a hard point under the fuselage or wing of an aircraft.

A tactical EW suite typically includes both ES and EA functions. The ES subsystem, such as a radar warning receiver (RWR), searches the electromagnetic environment in bands of interest, and then detects, sorts, characterizes, identifies, and locates radar emitters of interest. The ES operation is enhanced through the use of various databases that define emitter locations and characteristics. A threat library, for example, contains ranges of anticipated threat parameters such as frequency, pulse width, pulse repetition interval (PRI), and scan modulation that are used to identify specific threat emitter modes. The ES subsystem hands off detected intercepts identified as immediate threats to the EA subsystem, which then selects and executes the appropriate EA response against the radar. The EA system may continue, terminate, or alter the response based on a scripted procedure or a real-time assessment of technique effectiveness.

An EW system may vary its mode of operation over the course of an engagement. For example, an ingressing (flying into a region) SSJ aircraft may initially remain silent,

TABLE 12-1 ■ Radar and EW Frequency Band Designations

Radar Frequency Band Designations		EW Frequency Band Designations	
Band	Frequencies (MHz)	Band	Frequencies (MHz)
HF	3–30	A	0–250
VHF	30–300	B	250–500
UHF	300–1,000	C	500–1,000
L	1,000–2,000	D	1,000–2,000
S	2,000–4,000	E	2,000–3,000
C	4,000–8,000	F	3,000–4,000
X	8,000–12,000	G	4,000–6,000
Ku	12,000–18,000	H	6,000–8,000
K	18,000–27,000	I	8,000–10,000
Ka	27,000–40,000	J	10,000–20,000
V	40,000–75,000	K	20,000–40,000
W	75,000–110,000	L	40,000–60,000
		M	60,000–100,000

(Notes: HF, high frequency. VHF, very high frequency. UHF, ultra high frequency.)

using ES and mission profiles to maintain situation awareness. The aircraft might attempt to avoid detection through low-altitude flight and terrain masking. When the EW system determines that the radar is likely to detect the target, it might apply noise to prevent detection or false targets to delay or deny acquisition. Finally, it may employ noise or deceptive track gate pull-off techniques when it perceives that the radar has established track.

The EW community uses a different nomenclature for denoting frequency bands than the radar community. Table 12-1 summarizes the radar and EW band designations [5,8].

The words technique and waveform are used to describe specific EA approaches and implementations. There is ambiguity among practitioners regarding the precise definitions of these terms. For purposes of this chapter, a technique refers to a distinct EA concept or approach, whereas a waveform refers to a description of the specific temporal, spectral, or spatial characteristics of the EA signal being generated.

Another method of countering radar that is sometimes rolled into the EW category is the antiradiation missile (ARM). The ARM has a passive radio frequency (RF) seeker that provides guidance to the missile based on its track of the radar emission. The missile homes on the radar and disables or destroys it.

12.2.2 Jammer Types

There are two broad categories of jammer designs: noncoherent and coherent. Coherence in this context refers to the ability of a jammer to preserve or reproduce the phase characteristics of radar waveforms such as the intrapulse modulation of a pulse compression waveform or the interpulse phase relationship of a pulse-Doppler waveform. Within these categories are an innumerable variety of possible jammer designs. This chapter considers two simple, conceptual designs, represented in Figures 12-2 and 12-3. The key difference between the two categories is that the noncoherent jammer generates the EA signal through an independent waveform generator whereas the coherent jammer generates the EA signal

FIGURE 12-2 ■
Conceptual
noncoherent
jammer.

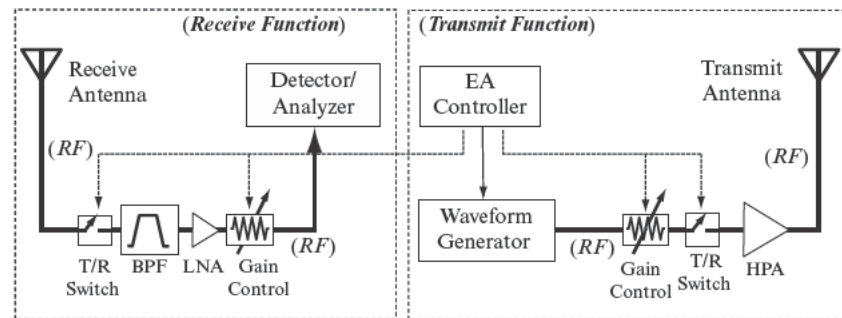
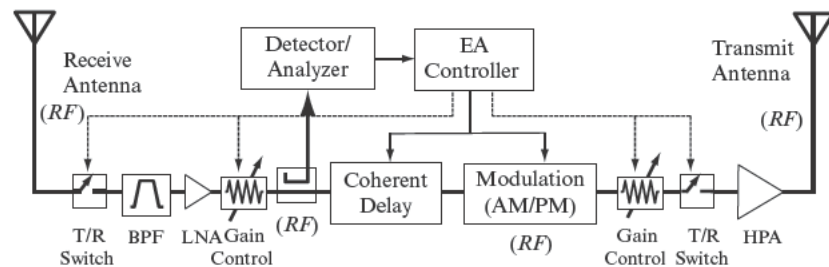


FIGURE 12-3 ■
Conceptual
coherent
jammer.



by modifying and retransmitting the actual received radar signal itself. In practice, actual systems may contain elements of both of these jammer types.

12.2.2.1 Noncoherent Jammers

Figure 12-2 is a conceptual diagram of a noncoherent jammer. Although such designs have been used to generate deceptive techniques against older, noncoherent radar systems, their role against modern, coherent radars is primarily limited to generating nondeceptive interference, and therefore they are commonly referred to simply as noise jammers. The type of jammer illustrated in the figure is sometimes referred to as a transponder: the jammer has its own signal generation capability but also responds in some manner—either temporally or spectrally—to the received signal [1,4].

Referring to the figure, the jammer contains a receive portion (left) and a transmit portion (right). Radar emissions are received through the receive antenna, passed through a RF front end, detected, and analyzed. Typical, generic RF front-end functions are represented by the transmit/receive (T/R) switch, band-pass filter (BPF), low-noise amplifier (LNA), and gain control attenuator; however, these are by no means standard in jammer designs. The detector/analyzer extracts frequency or timing information from the received signal and passes information to the EA controller in the transmit portion.

The EA controller commands the waveform generator to generate an appropriate EA waveform, which is then amplified by the high-power amplifier (HPA) and radiated out of the transmit antenna. The waveform generation may be accomplished by amplifying and band-limiting thermal noise, by frequency-modulating or phase-modulating a microwave oscillator, or through direct digital synthesis (DDS) of preprogrammed waveforms. Examples of HPAs are traveling-wave tube (TWT) amplifiers and klystron amplifiers. Generally it is preferable in EW systems to employ HPAs that offer 100% duty factor, or continuous wave (CW), operation. Other possible functions in the transmit chain depicted in the

figure include gain control and a second T/R switch, which may be an integrated part of the transmit amplifier.

The purpose of the T/R function in the jammer is to allow the receiver to sense the environment without being overpowered by the leakage from the transmitter back into the receiver. The leakage may occur internally within the system or externally through the antennas. The ratio of the peak HPA transmit power to the leakage input power at the receive antenna port is referred to as the system T/R isolation. If the leakage approaches or exceeds the detection threshold of the detector, the system will lose sensitivity. Generally, this type of jammer employs fairly high power, and T/R isolation is often not sufficient to maintain desired sensitivity while simultaneously transmitting and receiving. Under such conditions, the T/R switches are used to occasionally interrupt transmission to allow the receiver an uncontaminated sample of the RF environment. This process is referred to as look-through. During transmission, the receiver is blanked to prevent saturation or false detection; during reception, the transmitter is blanked to prevent high-power leakage into the receive path. The look-through periods are typically kept short relative to the transmitter-on times to maintain a high jamming duty cycle and, therefore, average power.

Some noise jammers exclude the receive function entirely from the design and simply generate noise in a manner that is independent of the current radar emissions. Such jammers forego the added cost of the receiver and the complexities of the transponder operation, but at the expense of being potentially less efficient in jamming the radar, as discussed in Section 12.2.3.1.

12.2.2.2 Coherent Jammers

Figure 12-3 is a conceptual diagram of a coherent jammer. The coherent jammer generates the EA signal by receiving, modifying, amplifying, and retransmitting the radar's own signal. A device that simply receives, amplifies, and retransmits the radar signal without any other change is providing a beacon to the radar—essentially an enhanced radar cross section (RCS) target—that provides little or no value in a support jamming role and may have disastrous consequences in a self-protection role. The modifications to the radar signal that produce the EA benefit are the coherent delay and the modulation functions shown in the figure.

The radar signal is received through the receive antenna, passes through an RF front end, and is then split, with one portion of the RF being routed to the detector/analyzer and the other continuing through the coherent delay, modulation, and transmitter chain and finally radiated through the transmit antenna. The RF front end and the transmitter chain are functionally the same as those described previously for the noncoherent transponder.

The detector function typically employs a diode detector and band-pass filter that rectifies and filters the received RF, essentially stripping off the carrier and leaving a magnitude-only pulse envelope. The output is sometimes referred to as detected video because the bandwidth needed to preserve the pulse envelope shape is comparable to that used to carry video information. The detected pulse envelope is often fed into a threshold detector, which generates a saturated (constant amplitude) video pulse, sometimes referred to as pulse-present indicator (PPIN), if the signal exceeds the threshold and zero voltage if it does not. The leading edge of the PPIN triggers a sample of a high-speed clock, which records the pulse time of arrival (TOA). The difference in time between the leading edge and trailing edge yields a pulse width (PW) measurement.

Some systems measure the frequency of each pulse with a device such as a frequency discriminator or instantaneous frequency measurement (IFM) receiver. An IFM receiver

splits the RF into parallel paths of different length and estimates the frequency based on the resulting phase differences between the paths. Depending on the receive antenna, some systems are able to measure angle of arrival (AOA) of each pulse. The AOA may be provided, for example, through amplitude comparison between a bank of fixed-beam, squinted antennas, phase comparison between different channels of an interferometer, the pointing angle of a steerable, directive antenna, or the peak output channel of an array, such as a Rotman lens. Generally frequency and AOA measurements are triggered by the leading edge of PPIN.

The single-pulse measurements such as TOA, PW, AOA, frequency, and amplitude, as applicable, are digitally encoded, packaged as a pulse-descriptor word (PDW), and sent to the analyzer for processing. Pulse TOAs from a series of pulses are used to determine the PRI of the radar pulse train, with the difference between successive pulse TOAs corresponding to the PRI. Often, the received pulses originate from multiple radar emitters, and the PDW data are used to first presort pulses based on similarity of PDW characteristics. If the PDW-based presorting still leaves intermixed emitter pulse data, a pulse-deinterleaving algorithm is invoked to separate the different pulse trains and measure their PRIs based on the TOA patterns.

The measured characteristics (e.g., frequency, PRI, PW) of the received signal are compared to a library of known radar parameters to make an identification of the intercepted emitter. If it corresponds to a radar mode that requires an EA response, the appropriate EA technique is selected and executed. The EA controller function determines the time-varying coherent delay or modulation to be applied to the received RF as well as T/R switching and gain control.

The modulation typically consists of amplitude modulation (AM) or phase modulation (PM). The modulation may be applied on the time scale of a radar pulse (intrapulse modulation), the radar PRI (interpulse modulation), or the radar antenna scan rate. The AM may be accomplished through on-off switching or through a variable attenuator. The PM may be implemented, for example, by varying a digitally controlled RF phase shifter or by applying a time-varying voltage to the helix of a TWT amplifier. In the latter case, an insertion phase is introduced in the TWT that is proportional to the helix voltage. A sawtooth-shaped control voltage can be used to linearly vary the insertion phase over a 360° extent, rapidly reset, and then repeat. The repetition frequency of the sawtooth modulation translates to an apparent frequency shift in the RF carrier—with each sweep of the ramp producing one 360° cycle. This method is often referred to as serrodyning [2,9]. The AM and PM are often used by a coherent jammer to create variable Doppler shifts.

The coherent delay, t_{delay} , produces an apparent range offset, R_{offset} , of

$$R_{\text{offset}} = \frac{ct_{\text{delay}}}{2} \quad (12.1)$$

where c is the speed of light. Thus, a $1\text{-}\mu\text{s}$ coherent delay corresponds to a range offset of 150 m, or approximately 500 ft, and a 100-ns delay corresponds an offset of 15 m, or approximately 50 ft.

Achieving a useful, time-variable, coherent delay, also referred to as coherent RF memory, presented a stubborn challenge to EW designers for many years. Waveguide and low-loss coaxial cabling, surface acoustic wave (SAW) delay lines, and more recently fiber optic delay lines, are all able to introduce coherent time delays in the received signal, but they have limited applicability due to signal degradation or lack of variability. Ideally,

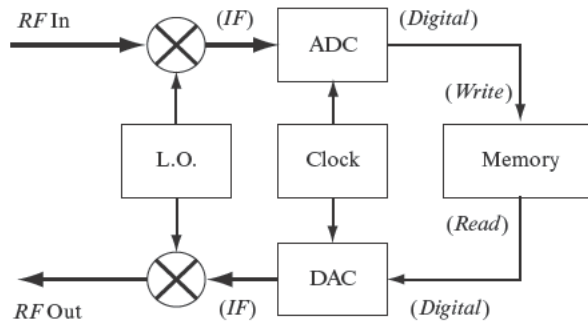


FIGURE 12-4 ■
Digital RF memory
concept.

coherent delays should be variable between nominally zero delay and approximately the PRI of the radar in step sizes less than the range resolution of the radar. With radar PRIs as high as hundreds of microseconds and resolutions as low of fractions of a microsecond, achieving both the requisite delay extent and variability has proved impractical through the above technologies.

The arrival of a technology known as digital RF memory (DRFM) proved to be a huge boon to the implementation of coherent delays. The DRFM, conceptually depicted in Figure 12-4, digitizes the input signal with a high-speed analog-to-digital converter (ADC) and sequentially writes the digital samples into memory. After a desired time delay, it then reads the data from the appropriate memory locations and converts them back to analog voltages with a digital-to-analog converter (DAC). The input signal must be at a frequency and contained within a bandwidth that can successfully be sampled by the ADC. Typically, this requires frequency translation of the received RF signal to an intermediate frequency (IF) prior to the ADC and then translation back from IF to RF following the DAC, as illustrated in the figure. For example, if the ADC can sample at a 50-MHz rate, then the input signal waveform must be contained within a bandwidth of 25 MHz or less to satisfy the Nyquist sampling criterion. Alternatively, a dual-channel in-phase/quadrature (I/Q) configuration can be used to achieve the full sampling bandwidth. A common, stable local oscillator (LO) is used for both down-conversion and up-conversion stages, thereby ensuring that the output RF is phase-coherent with the input RF. Because the signal does not degrade once it is in digital form, and because the time delay between writing the data to and reading them from memory is under programmable control, the DRFM is capable of achieving arbitrarily long, coherent time delays, with variability on the order of the ADC and DAC sampling interval.

As with other ADC and DAC applications, trade-offs are often required in a DRFM design between sampling rate, or, equivalently, instantaneous bandwidth (IBW), and the number of bits, or resolution. The wide IBW helps capture wideband radar waveforms, such as high-range resolution, pulse compression waveforms, as well as radars with frequency hopping capability. The high resolution helps provide improved dynamic range, with each bit of resolution on the ADC or DAC being equivalent to a factor of two in voltage range, or 6 dB in power (power is proportional to voltage squared, and $10 \log 2^2 = 6$).

The DRFM is often used in one of two modes, commonly referred to as pipeline and stored delay. In pipeline mode, the input signal is continually sampled and replayed with a specific delay. The DRFM can read from memory while it is writing to memory, allowing simultaneous T/R operation in a jammer if desired. The coherent delay—the difference in time between writing and reading data—is varied with time according to the desired

EA waveform. In stored delay mode, the input signal must be detected, and all or part of the received pulse must first be stored in memory before any of the digitized samples can be read out. This mode, therefore, does not lend itself to simultaneously receive (writing) and transmit (reading) operation.

Coherent jammers tend to use broadband, CW amplifiers. This allows them to operate against multiple emitters distributed across the RF spectrum, timeshare between multiple emitters simultaneously, and better accommodate the high duty factor of coherent radar waveforms. Common technologies are the TWT amplifier, solid-state amplifier (SSA), and the microwave power module (MPM) amplifier. The MPM uses an SSA for the initial gain stage and a TWT for the final-gain and high-power stage. The SSA offers smaller size and better noise figure, while the TWT offers higher peak power. Typical HPA power levels of interest for coherent EA range from a few watts to hundreds of watts, depending on the application.

The architecture in Figure 12-3 is actually a blend between a traditional coherent repeater and what is sometimes referred to as a coherent, or quasi-coherent, transponder [1,2,4,10]. In a classic coherent repeater, the coherent delay is very minimal, achieved through a length of microwave transmission line for example, and the signal remains at RF throughout its transit through the jammer. A repeater that can simultaneously receive and transmit is sometimes referred to as a straight-through repeater or modulated straight-through repeater. A repeater that cannot simultaneously receive and transmit due to isolation limitations will employ the short coherent RF delay to capture a segment of the incoming pulse and rapidly alternate between receive and transmit states. This type of repeater is referred to as a chopped repeater. The addition of a DRFM, or other coherent memory, which can provide very long coherent delays and can continually reproduce the sampled signal for time periods well beyond the received pulse width, allows the jammer to operate in a transponder-type mode. In the diagram of Figure 12-3, bypassing the coherent delay would allow the jammer to operate as a traditional coherent straight-through repeater or chopped repeater. For convenience, we will forego the previous distinction in favor of the common vernacular and simply refer to the hybrid architecture of the figure as a coherent repeater.

12.2.3 Masking EA Techniques

Masking EA attempts to prevent target detection or deny measurement of target position or velocity. Masking can be achieved through either the noncoherent or coherent jammer designs of Figures 12-2 and 12-3. The noncoherent jammer generates noise over a bandwidth that is comparable to or exceeds the radar waveform bandwidth. For example if the radar employs a 10 MHz pulse compression waveform, the jammer would be designed to generate a noise spectrum that is at least 10 MHz wide against that radar. The coherent jammer can generate such noise as well but also has the ability to focus its energy much more efficiently in the range and Doppler domains, making it better matched to the radar detector.

12.2.3.1 Noncoherent Masking Techniques

The noncoherent jammer of Figure 12-2 can generate noise in either a transmit-only, noise generator mode or a responsive, noise transponder mode [1,2]. The noise generator mode is used to generate techniques such as spot noise (SN), barrage noise (BN), and blinking

variations of each—blinking spot noise (BSN) and blinking barrage noise (BBN). The transponder mode generates techniques such as responsive spot noise (RSN) and noise cover pulse (NCP).

One method of generating noise waveforms is by directly amplifying and band-limiting the thermal noise occurring naturally in electronic components [2]. This produces a Gaussian distribution of the noise, which is appealing because it closely matches the thermal noise in a radar receiver. The disadvantage of the Gaussian noise is that it implicitly requires AM of the EA signal. This results in the average output power of the transmitter being significantly less than its peak output power rating, a fairly expensive impact given the high cost of HPAs. An alternate approach is to apply either frequency modulation (FM) or rapid PM to an RF source. This allows the amplifier to be continuously operated at peak output power as the spectral spreading is achieved entirely without resorting to AM.

Spot noise produces a narrowband noise waveform that is intended to cover the radar IF bandwidth. Typical SN bandwidths are several MHz to several tens of MHz. This allows the jamming energy to be concentrated at the radar frequency. Spot noise is used when the radar frequency is known in advance. Multiple spot noise is a variant in which multiple frequencies are covered with separate SN waveforms.

Barrage noise is used when the specific operating frequency of a victim radar is not known in advance or if the jammer must simultaneously cover multiple radar emitters in the environment distributed over a wide spectral bandwidth. The BN energy is spread over a much wider spectrum than SN, possibly hundreds of MHz, for example. As a result, the jamming energy is diluted relative to any individual radar being covered. A BN jammer may achieve its wideband coverage by rapidly sweeping or stepping a SN waveform, or even a CW tone, across the RF band. Blinking noise—blinking spot noise or blinking barrage noise—is sometimes used instead of continuous noise. One reason is to deprive an RF homing missile of a continuous signal to track in a home-on-jam mode, described in Section 12.10.6.

Responsive spot noise and NCP are transponder techniques in which the jammer receives the radar transmission and generates a response that is somehow dependent on the input. The RSN is accomplished by the jammer determining the frequency of the radar signal and rapidly setting-on a SN waveform at that frequency. The jammer maintains continuous transmission except for occasional, brief look-through interruptions, during which the frequency or timing is updated as necessary. The RSN allows the jammer to cover a wide RF bandwidth more effectively than a BN approach because it concentrates its noise at the radar's current frequency. For example, if the RSN bandwidth is 20 MHz, and the BN bandwidth is 200 MHz, the RSN jammer can achieve nominally ten times the in-band power of a BN jammer. Key issues for RSN implementation are frequency measurement, set-on accuracy, and the look-through update function.

The NCP technique is primarily a SSJ technique used to deny range against pulse compression waveforms. The jammer generates a pulse of noise that is triggered by and concurrent with the received pulse. The noise, which may be comparable in bandwidth to SN or BN, is strong enough to mask the target return even after the radar performs its pulse compression. This creates an extended-range return spread over many range bins that hides the true target range. Often this is referred to simply as cover pulse, but we use the name NCP in this chapter to explicitly distinguish it from a coherent cover pulse, which may occupy only a single range bin or a few range bins and is not necessarily restricted to pulse compression waveforms.

12.2.3.2 Coherent Masking Techniques

A coherent jammer such as the one represented in Figure 12-3 can potentially create masking that is much more focused and more insidious than the noncoherent jammer techniques. Coherent masking is sometimes referred to as coherent noise, a term that annoys some purists but is actually fairly descriptive of what occurs.

One form of coherent masking is Doppler noise (DN), also referred to as narrowband Doppler noise (NBDN) or narrowband repeater noise (NBRN) [6]. A coherent repeater achieves DN by applying a PM (or, less commonly, AM) to the received signal. Instead of the return focusing up in the Doppler filter corresponding to the jamming platform, it is spread across multiple Doppler filters. Typical DN bandwidths range from a few kHz to several tens of kHz—orders of magnitude smaller than RSN-type bandwidths. An analogous approach in the range dimension uses a DRFM or other coherent memory to capture and continuously replay the received radar signal across multiple range bins or the entire PRI in a CW range denial mode. This can be combined with DN to produce extended range DN.

Range bin masking and velocity bin masking, or Doppler bin masking, are similar to the preceding techniques but are distinguished from them in this chapter in that bin masking is restricted to a smaller region in the range and Doppler domains, with a specific objective of attacking the radar constant false alarm rate (CFAR) detector. (This distinction is not universal.) The jammer attempts to place coherent returns in range or Doppler bins that are used by the radar to form its noise estimate for setting the CFAR detection threshold. The EA biases the noise estimate higher, potentially causing the threshold to be raised enough to prevent the target from being detected. The range or velocity bin masking energy is typically spread over a contiguous range or Doppler extent; however, similar effects can be achieved if the energy is focused up in distinct range-Doppler bins, such as would occur for multiple targets returns.

Another form of coherent masking is automatic gain control (AGC) capture in which the jammer exploits the radar AGC to reduce sensitivity, as described in Section 12.8.3.

12.2.4 Deception EA Techniques

Deception EA techniques inject target-like returns in the radar that contain false information. The primary classes of techniques are false targets and track deception/denial. Noncoherent jammers were once used to generate deception techniques against older, pulsed radar systems; however, against modern, coherent radars, this role has been almost exclusively fulfilled by the coherent jammer.

12.2.4.1 False Targets

False targets are used to create distraction or confusion, primarily against search and acquisition modes, and can be generated in range, velocity, or angle. Range false targets (RFTs) typically require use of a DRFM or other coherent RF memory to create range-delayed, coherent copies of the radar pulse. Against pulse-Doppler radar, a DRFM can be used to exploit the range ambiguity of the radar waveform and make the false target appear to be in front of the actual jammer position in range. This is accomplished by delaying the received signal by one PRI minus the desired time advance. The jammer continually monitors the PRI using an ES function referred to as a PRI tracker and updates this delay as needed. Velocity false targets (VFT), or Doppler false targets, can be superimposed on RFTs or generated independently through a straight-through repeater by AM or PM

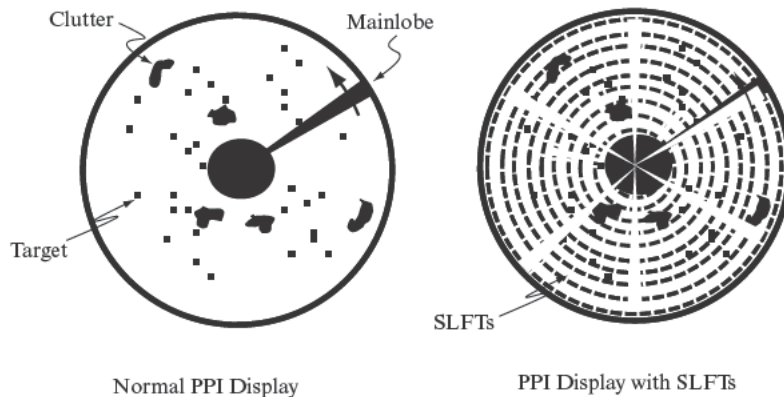


FIGURE 12-5 ■
Effect of sidelobe
false targets on
search radar.

of the received signal. To make the false targets appear credible against pulse-Doppler radar, the RFTs and VFTs must be coordinated to appear kinematically consistent: if the RFT is moved with a range rate of v , the associated VFT must have a Doppler shift of $2v/\lambda$, where λ is the radar wavelength, or equivalently $2vf_c/c$, where f_c is the radar frequency.

Angle false targets, here referring specifically to sidelobe false targets (SLFTs), are range or velocity false targets generated in the radar sidelobes. The jammer must have sufficient sensitivity and power to overcome the radar sidelobe pattern attenuation relative to the main lobe. The conceptual representation of the effect on a plan position indicator (PPI) display of a surveillance radar is illustrated in Figure 12-5. The SLFTs make it difficult to determine the true target among the many false ones and may potentially overload the radar processor with false detections. If the jammer is able to synchronize the SLFTs with the radar antenna scan pattern, such as with a circularly rotating radar antenna, the false targets can be made to appear consistently at the same position from scan to scan, improving their credibility.

12.2.4.2 Range and Velocity Track Deception

The objective of range or velocity track deception is to draw the radar track gate away from the true range or velocity of the target to some other desired position. A tracking radar may have track filters in range, Doppler, or both dimensions. The radar attempts to keep the track gate centered on the target based on the power centroid in the track gate. As the target begins to slip out of the gate, an error signal is generated to move the track gate back on the target. A coherent SSJ generates a return that is initially in the same range-Doppler cell as the target and much stronger in amplitude than the target return itself, sometimes referred to as the aircraft skin return. The jammer then varies the range or Doppler of its false target away from the cell containing the target. The radar's track error signal is dominated by the jammer because of its higher power, and the track gate is thereby drawn away from the true target return. This type of EA is sometimes called a seduction technique.

The range track deception technique is variously referred to as range gate capture (RGC), range gate pull-off (RGPO), range gate walk-off (RGWO), or range gate stealing (RGS) [2,6,9,10]. Similarly the velocity technique is referred to as velocity gate capture, velocity gate pull-off (VGPO), velocity gate walk-off (VGWO), or velocity gate stealing (VGS). These are essentially different names for the same idea. Sometimes the terms range

gate pull-in (RGPI) and velocity gate pull-in (VGPI) are used to explicitly distinguish the direction of the track gate motion, but this distinction is often ignored.

Range gate pull-off and VGPO are said to be coordinated when the two techniques are executed in a manner that produces a kinematically consistent false target in both range and Doppler dimensions, as described in Section 12.2.4.1 for RFTs and VFTs. A simple coordinated pull-off might use a constant acceleration, a , in which the range is varied parabolically with time ($R = at^2/2$), and the velocity is varied linearly with time ($v = at$).

The pull-off is sometimes defined in four phases: dwell, walk, hold, and drop. During the dwell, the jammer turns on with a strong signal in the same cell as the target return. During the walk, the jammer draws the track gate away from the target. During the hold, the jammer stops moving the track gate and pauses at the current location. During the drop, the jammer turns off. Any of the phases except the walk phase may be omitted in the sequence. The pull-off sequence is repeated as needed.

The range and velocity track gate stealing techniques might be used for several objectives: to induce large track errors, thereby degrading the radar track quality; to break the radar's track, forcing time-consuming reacquisition; to transfer the radar track to another return such as clutter or chaff; and to facilitate the execution of EA techniques, such as angle deception, that require a very high jam-to-signal ratio (JSR) (with no target signal in the denominator of the ratio, the JSR becomes infinite).

12.2.4.3 Angle Track Deception

Angle track deception techniques are used by SSJs to create angle track errors or angle break-locks (track breaks) against radars supporting weapon system engagement, such as fire-control radars or RF missile seekers. These fall into two broad categories: on-board, in which the EA signal originates from the target itself; and off-board, in which the EA signal originates from an object spatially separated from the target being protected [1,2,7].

Examples of on-board techniques include inverse gain (IG), swept AM (SAM), AGC deception, image jamming, cross-polarization (XPOL), cross-eye (XEYE), terrain bounce (TB), and jammed chaff (JAFF). Inverse gain and SAM, described in Section 12.5.8, are limited to conical scan type angle tracking radars. The AGC deception and image jamming techniques, described in Sections 12.8.3 and 12.8.2, exploit limitations in the radar receiver dynamic range and image rejection, respectively. Cross-polarization, commonly referred to as cross-pol, and XEYE produce distorted responses of a monopulse angle tracking radar. Cross-pol accomplishes this by generating a signal that is orthogonally polarized to the radar antenna; XEYE accomplishes this by generating a spatial interference pattern from two widely separated antennas, causing a warped, or distorted, phase front across the face of the antenna. Terrain bounce is achieved by an SSJ that flies at low altitude and illuminates the terrain. If the reflection from the terrain is stronger than the target return and any direct-path leakage from the jammer, a homing missile will track the reflection and fly into the terrain. Jammed chaff is an airborne version of TB: the jamming platform illuminates a chaff cloud with sufficient power to create a spatially displaced target for a homing missile. Both TB and JAFF require directional antennas that minimize direct-path radiation to the radar. All of the on-board techniques require very high JSR to be effective.

Examples of off-board techniques are chaff, expendable decoy, towed RF decoy (TRD), and cooperative blinking noise (CBN). Chaff bursts consist of huge quantities of small reflective material such as wavelength-sized, metal-coated glass filaments, which are ejected from an aircraft and then bloom into a large reflective, hovering cloud that

provides a return at a different angle from the dispensing aircraft. (Other applications include much larger regions of coverage called corridor chaff or different types of chaff, such as absorptive.) An expendable decoy is ejected from a target aircraft and provides an alternate angle target for a RF homing missile. Both active (radiating) and passive (reflecting) decoys are used. A TRD is an active decoy that is towed behind an SSJ aircraft at sufficient distance that it will draw a homing missile away from the aircraft toward the decoy. Cooperative blinking noise is a mutual-protection technique in which two aircraft employ blinking noise to create angle errors in the radar. Both platforms are in the main beam but separated by a significant fraction of the radar beamwidth. As they blink on and off the radar receives mixed angle error signals, leading to an unstable track.

As is evident from the preceding summary, there is a wide range of possible angle deception techniques, and a proper description of them is well beyond the scope of this chapter. Some of them will be addressed in more detail in subsequent sections, but interested readers are advised to follow up with descriptions in [1,2,6,7,9].

12.3 | EW-RELATED FORMULAS

The assessment of the potential effectiveness of an EA or EP technique often requires estimating the strength of the jamming signal relative to that of the target and thermal noise as well as the power received by the jammer from the radar. This section presents formulas that provide a starting point for such analyses. Of particular interest are the JSR and the jam-to-noise ratio (JNR). Section 12.1.3 provides definitions for the symbols used in the following sections.

12.3.1 Signal and Thermal Noise Formulas

From the derivation of the radar range equation presented in [8], we can represent the peak power, P_r , of a target return received through the antenna of a monostatic radar as

$$P_r = \frac{P_t G_t G_r \lambda^2 \sigma}{L_s (4\pi)^3 R^4} \quad (12.2)$$

The actual signal, S , available for detection receives the additional processing gains of the time-bandwidth product, τB , if pulse compression is used, and the number of pulses coherently integrated, n , if a multiple pulse coherent waveform, such as pulse-Doppler, is used:

$$S = \frac{P_t G_t G_r \lambda^2 \sigma n \tau B}{L_s (4\pi)^3 R^4} \quad (12.3)$$

Equation 12.3 is also valid for cases in which neither pulse compression nor a multiple-pulse waveform is used: when pulse compression is not used, the signal bandwidth is approximately equal to the inverse of the pulse width, making their product equal to unity; and when only a single-pulse waveform is used, then $n = 1$ as well.

Many EW engineers like to use the so-called link equation approach in writing Equation 12.3 for the target signal [2,11]. This approach rearranges parts of the equation into several dimensionless terms, allowing the resulting formula to be interpreted as an RF

circuit, with a power source and gain and loss stages:

$$S = P_t G_t \left(\frac{\lambda}{4\pi R} \right)^2 \left(\frac{4\pi\sigma}{\lambda^2} \right) \left(\frac{\lambda}{4\pi R} \right)^2 G_r \left(\frac{n\tau B}{L_s} \right) \quad (12.4)$$

The term $(\lambda/(4\pi R))^2$ is referred to as free-space attenuation, or free-space loss, which accounts for the range-dependent power variation each way. The term $(4\pi\sigma/\lambda^2)$ is referred to the target gain, reminiscent of the formula that converts antenna area, A , into antenna gain $(4\pi A/\lambda^2)$. Thus, reading through the terms from left to right, the radar transmits a signal at a peak power P_t ; the transmit antenna provides gain G_t ; the radiated signal experiences a free-space loss going to the target; the target has a gain reflecting the signal back to the radar; the reflected signal suffers the same free-space loss traveling back to the radar; the radar receive antenna provides gain G_r ; and finally, there are signal processing gains and system losses, yielding the final signal magnitude in units of power.

The competing thermal noise power in the receiver bandwidth is $N = kTFB$, where F is the radar receive noise figure. The familiar signal-to-noise ratio (SNR) of the radar range equation is thus

$$SNR = \frac{P_t G_t G_r \lambda^2 \sigma n \tau}{L_s (4\pi)^3 R^4 k T F} \quad (12.5)$$

We have been expressing the signal and noise levels in units of power. Strictly speaking, the pulse compression and multiple-pulse signal processing gains do not actually increase the signal power in an absolute sense but rather convey an increase in signal relative to the noise in the matched filter detector. Stated differently, the signal level of Equation 12.3 can be thought of as the equivalent peak received power if the radar did not have subsequent signal processing gain.

An alternate way of interpreting the radar range equation is to think in terms of energy available to the detector. The signal energy is the peak power multiplied by time or, more precisely, the cumulative time of the return signal used for detection. The cumulative time is the number of pulses multiplied by the duration of each pulse. Thus, the signal energy becomes

$$S = P_t n \tau = \frac{P_t G_t G_r \lambda^2 \sigma n \tau}{L_s (4\pi)^3 R^4} \quad (12.6)$$

The thermal noise energy that competes with the signal in the matched filter detector is $N = kTF$, which is equal to the noise power spectral density. This approach may provide a more accurate description of the detection process; however, most people tend to think in terms of power, so we will continue using that perspective.

12.3.2 Keeping Track of Losses

As described in [8] radar system losses as represented in the term L_s may be quite significant in determining target SNR. Such losses are not in general the same for the jammer as they are for the target, however. First, jammers such as SOJs and SIJs are likely to be at significantly different ranges, altitudes, and azimuths than the target, incurring different propagation losses. Second, the transmit power of some jammers is independent of received power, in which case any transmit losses and propagation losses to the jammer are irrelevant. And third, a noise jammer and some coherent repeater modes will not

incur the same straddle and signal processing losses as the target return. It is certainly important to keep track of these different losses in actual practice; however, for the sake of simplicity, we will omit the radar loss terms for the target and jammer in the following discussions.

12.3.3 Noise Jammer Formulas

The in-band power of a noise jammer is derived in a manner similar to the signal power in the radar range equation. In deriving general expressions, we must bear in mind that the jammer is not necessarily co-located with the target, an example being an SOJ. Thus, we must explicitly distinguish the range to the target from the range to the jammer and the antenna pattern gains toward the target from those toward the jammer. With that in mind, the noise jamming power is given by

$$\begin{aligned} J &= (P_j G_{j,tx}) \left(\frac{1}{4\pi R_j^2} \right) \left(\frac{\lambda^2 G_{rj}}{4\pi} \right) \left(\frac{B}{B_j} \right) d_j \\ &= \frac{P_j G_{j,tx} \lambda^2 G_{rj} B d_j}{(4\pi)^2 R_j^2 B_j} \end{aligned} \quad (12.7)$$

The parenthesized terms in the top line of the above equation are, from left to right, the jammer effective radiate power (ERP) at the radar polarization (transmit power (P_j) times antenna gain including losses ($G_{j,tx}$)); the spherical spreading of the jamming signal at the radar antenna at range R_j ; the capture area of the radar antenna whose receive gain toward the jammer is G_{rj} ; the fraction of jamming noise spectrum (B_j) that is within the radar waveform bandwidth (B); and the jammer transmit duty factor, d_j , which converts peak power to average power if the jammer is not on continuously. The jammer transmit antenna gain, $G_{j,tx}$, is the effective antenna gain at the receive polarization of the radar antenna. The duty factor is especially important for fast-blinking noise waveforms but also accounts for transmitter off-times employed by a transponder during look-through intervals. The jammer noise bandwidth term, B_j , is defined to be greater than or equal to the radar waveform bandwidth, as the noncoherent jammer signal is ultimately decorrelated by the radar matched filter regardless of its bandwidth.

The JNR for a noise jammer is obtained by dividing Equation 12.7 by the noise power, $kTfB$:

$$JNR = \frac{P_j G_{j,tx} \lambda^2 G_{rj} d_j}{(4\pi)^2 R_j^2 B_j kTf} \quad (12.8)$$

The JNR gives insight into the degradation in radar sensitivity caused by the noise jammer. When the jammer power is comparable to the thermal noise power, a more accurate representation of the relative degradation is the ratio is $(J + N)/N$.

The JSR for a noise jammer is obtained by dividing Equation 12.7 by Equation 12.3 (where we ignore the loss term, L_s , as per the discussion in Section 12.3.2):

$$JSR = \frac{P_j G_{j,tx} G_{rj} d_j 4\pi R^4}{R_j^2 B_j P_i G_r \sigma n \tau} \quad (12.9)$$

Inspection of Equation 12.9 immediately suggests several radar design parameters that potentially affect the radar's vulnerability to noise jamming: ERP, receive antenna gain

TABLE 12-2 ■ Noise Jammer Example Parameters

Radar Parameter	Value	Jammer Parameter	Value
Transmit power (P_t)	5000 W	ESJ transmit power (P_j)	100 W
Transmit antenna gain (G_t)	30 dBi	ESJ transmit antenna gain at radar receive polarization ($G_{j,tx}$)	3 dBi
Receive antenna gain (G_r)	30 dBi	ESJ receive antenna gain at radar transmit polarization ($G_{j,rx}$)	3 dBi
Transmit antenna gain to sidelobe jammer (G_{tj})	-5 dBi	ESJ noise bandwidth (B_j)	100 MHz
Receive antenna gain to sidelobe jammer (G_{rj})	-5 dBi	ESJ duty factor (d_j)	1.0
Frequency (f_c)	6 GHz	SOJ transmit power (P_j)	500 W
Radar wavelength (λ)	0.05 m	SOJ transmit antenna gain at radar receive polarization ($G_{j,tx}$)	17 dBi
Pulse width (τ)	2 μ s	SOJ noise bandwidth (B_j)	100 MHz
Waveform bandwidth (B)	10 MHz	SOJ duty factor (d_j)	1.0
Number of pulses (n)	16		
System noise figure (F)	5 dB		

toward the target, receive antenna gain toward the jammer, waveform pulse width, and number of pulses coherently integrated. The equation also indicates that the target return increases faster with decreasing range (R^{-4}) than the jammer return (R_j^{-2}).

For an SSJ, and generally also for an ESJ, the target and jammer are both at the same range and angle, making the JSR

$$JSR_{SSJ} = \frac{P_j G_j d_j 4\pi R^2}{B_j P_t G_t \sigma n \tau} \quad (12.10)$$

Note that the JSR is directly proportional to range squared for the SSJ (and typically the ESJ). Thus as the target gets closer, the JSR decreases, and eventually the signal will become stronger than the jammer.

12.3.4 Noise Jammer Computation Example

To illustrate the application of the noise jammer formulas, consider a ground-based radar that is trying to detect an incoming, airborne target whose RCS is 10 m². An ESJ accompanies the target at the same range and angle, and an SOJ is located 100 km from the radar and is in the radar sidelobe; both jammers use barrage noise. The assumed parameters for the radar and jammers are summarized in Table 12-2. The ESJ ERP is 23 dBW, and the SOJ ERP is 44 dBW; otherwise, both jammers have the same noise bandwidth and duty factor.

Using the assumed values in Table 12-2, Figure 12-6 plots the target signal, thermal noise, and jammer noise as a function of the target range. The plot uses a logarithmic scale for range and plots the various parameters in units of power (dBW). The received target signal (S) varies as R^{-4} , increasing with decreasing range. The ESJ flies at the same range as the target, and therefore its noise level increases with decreasing range, but its power level varies as R^{-2} because it experiences the R^2 spreading loss only in one direction. The SOJ is at a constant 100-km range, and therefore its noise level is independent of the target range.

If we assume that the radar requires 15-dB SNR for detection, then in the absence of any jamming, the detection range would be approximately 50 km: this is the point at

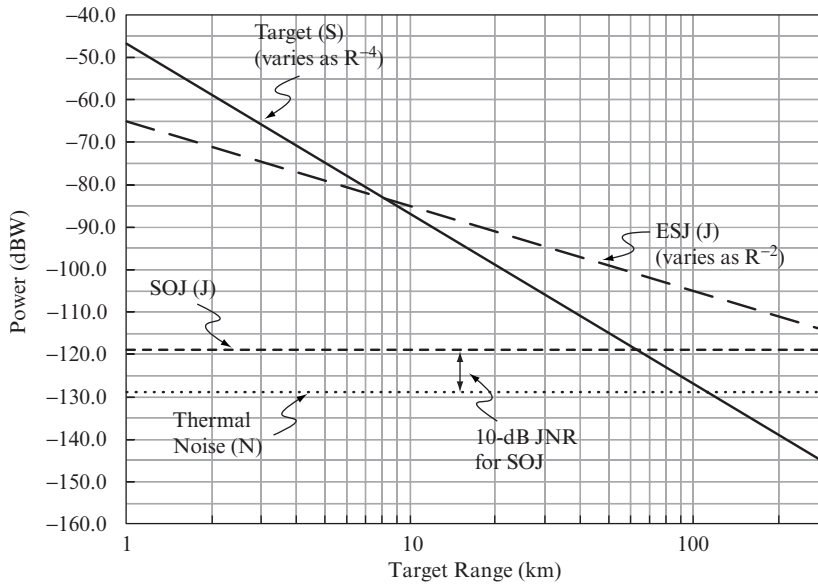


FIGURE 12-6 ■
Noise jamming
power example.

which the S curve exceeds the N curve by 15 dB. The sidelobe SOJ produces a 10-dB JNR and decreases the detection range—the point at which the signal exceeds the noise by 15 dB—to approximately 30 km. The ESJ, despite having 21 dB less ERP than the SOJ, is able to screen the target at much closer range. The target begins to break through at approximately 8 km but does not satisfy the 15-dB detection criterion until almost 1 km, at which point the radar's detection report may not have much value. The specific values used in the example are not as important as understanding the basic approach and trends of the different parameters.

12.3.5 Coherent Jammer Formulas

The coherent jammer can be thought of as operating in either a constant power mode, in which the transmitted power is independent of the received power, or a constant gain mode in which the transmitted power is directly proportional to the received power. Jammers operating in these modes are sometimes referred to as a saturated repeater and a linear repeater, respectively. Strictly speaking, this terminology is ambiguous because in theory a jammer can use AGC to produce a constant-power response without saturating any of its components or to produce a constant-gain response without all of its components operating linearly. Despite this technicality, the linear/saturated terminology is a convenient choice.

The formula for a saturated repeater is derived in a manner similar to the target signal and noise jammer formulas of the previous sections. Using the link-equation approach of Equation 12.4, the saturated repeater jammer level, J_{SAT} , in a radar detection cell is

$$\begin{aligned}
 J_{SAT} &= (P_j G_{j,tx}) \left(\frac{\lambda}{4\pi R_j} \right)^2 G_{rj} (n\tau B) \left(\frac{d_j}{n_j} \right) \\
 &= \frac{P_j G_{j,tx} \lambda^2 G_{rj} n\tau B d_j}{(4\pi)^2 R_j^2 n_j}
 \end{aligned} \tag{12.11}$$

This resembles Equation 12.7 for the noise jammer except for the last parenthesized terms, which reflect the fact that the jammer is coherent and can recover some or all of the radar coherent processing gain, $n\tau B$. The duty factor term for the coherent jammer, d_j , is not the actual transmit duty factor but rather the fraction of the radar pulse train that is repeated back to the radar. Thus if the jammer responds only on alternate pulses, $d_j = 0.5$. The term n_j in the denominator is the number of range or Doppler cells across which the jammer simultaneously spreads its power. For example, if the jammer puts all of its power into one false target occupying a single range-Doppler cell, then $n_j = 1$. If the jammer spreads its power over 10 Doppler filters in a single range cell using DN, then $n_j = 10$. If the jammer spreads its energy over five range cells all within the uncompressed pulse width of a pulse compression waveform, then $n_j = 5$. If the latter two conditions are combined, then $n_j = 10 \times 5 = 50$. Note that if the range spacing of the jammer returns exceeds the radar pulse width, the jammer incurs no penalty in power, as the returns do not simultaneously compete with each other for the HPA's available power.

The formula for a linear repeater is similar, but we must first begin at the radar transmission because the output power of the jammer is proportional to the input power. We can think of the constant-gain repeater as having some internal, electronic small-signal gain, $g_{j,el}$, between its receive antenna port and its transmit antenna port, and some maximum available peak power, P_j , in its HPA. As the received signal increases, the output of the transmitter increases by the small-signal gain $g_{j,el}$, until eventually the amplified signal reaches the peak output power of the HPA. As the received signal increases beyond this point, the HPA operates in compression, and the jammer reverts to the constant power repeater. The linear repeater jammer level in the constant-gain region, J_{LIN} , is (invoking the link equation approach)

$$\begin{aligned}
 J_{LIN} &= \left[(P_t G_{tj}) \left(\frac{\lambda}{4\pi R_j} \right)^2 G_{j,rx} \right] g_{j,el} G_{j,tx} \left(\frac{\lambda}{4\pi R_j} \right)^2 G_{rj} (n\tau B) \left(\frac{d_j}{n_j} \right) \\
 &= \frac{P_t G_{tj} G_{rj} \lambda^4 G_{j,rx} g_{j,el} G_{j,tx} n\tau B d_j}{(4\pi)^4 R_j^4 n_j}
 \end{aligned} \tag{12.12}$$

The power received by the jammer at its receive antenna port corresponds to the bracketed set of terms in the top line of the equation that begins with the radar transmit power, P_t , and ends with the term $G_{j,rx}$ —the effective jammer receive antenna gain, including losses, at the radar transmit polarization. When we then include the electronic gain $g_{j,el}$ with the bracketed terms, the result is the output power of the jammer transmitter. From there on, the equation is the same as for the constant-power repeater. The jammer power from a linear repeater is actually the lesser of either J_{SAT} or J_{LIN} in Equations 12.11 and 12.12, respectively, as the jammer output power is ultimately limited by the maximum power available through its HPA.

The expressions for JNR for the two repeater modes are obtained by dividing J_{SAT} and J_{LIN} by $kTFB$. The JNR for the coherent jammers has a similar connotation to that of a noise jammer when DN is used, but for false target techniques the JNR is an indication of whether the jammer false target is strong enough to be detected by the radar.

The expressions for JSR for the two repeater modes are obtained by dividing J_{SAT} and J_{LIN} by the target signal, S , of Equation 12.6 (ignoring the loss term, L_s , as per

Section 12.3.2):

$$JSR_{SAT} = \frac{P_j G_{j,tx} G_{rj} d_j 4\pi R^4}{R_j^2 P_t G_t G_r \sigma n_j} \quad (12.13)$$

$$JSR_{LIN} = \frac{G_{tj} G_{rj} G_{j,rx} G_{j,tx} g_{j,el} \lambda^2 R^4 d_j}{G_t G_r 4\pi \sigma R_j^4 n_j} \quad (12.14)$$

These expressions are general enough to apply to SSJ, ESJ, SIJ, and (with some exceptions) SOJ geometries. For the specific case of an SSJ, as well as for the case of an ESJ that is in close proximity to the target, the previous formulas simplify somewhat, as we set $G_{tj} = G_t$, $G_{rj} = G_r$, and $R_j = R$. This yields expressions for the saturated SSJ repeater JSR and the linear SSJ repeater JSR as, respectively,

$$JSR_{SAT,SSJ} = \frac{P_j G_{j,tx} 4\pi R^2 d_j}{P_t G_t \sigma n_j} \quad (12.15)$$

$$JSR_{LIN,SSJ} = \frac{(G_{j,rx} G_{j,tx} g_{j,el})}{(4\pi \sigma / \lambda^2)} \left(\frac{d_j}{n_j} \right) \quad (12.16)$$

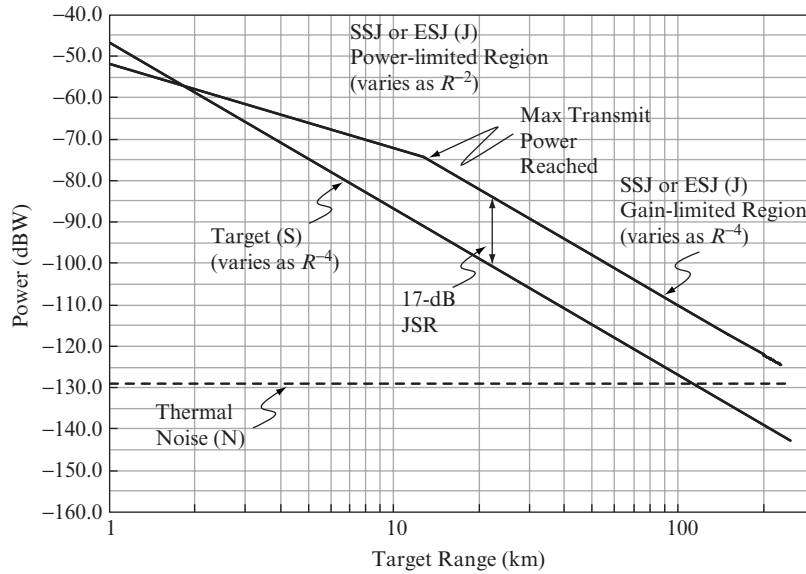
The parenthesized terms in the latter expression provide insight into how to interpret the JSR for the constant-gain repeater. The term $(G_{j,rx} G_{j,tx} g_{j,el})$ is the net repeater throughput gain at the radar polarization: antenna gain minus losses plus electronic gain (speaking in dB). The term $(4\pi \sigma / \lambda^2)$ is the so-called target gain from the link equation, Equation 12.4. When these two terms are equal, the SSJ produces a signal level equal to that of a target of RCS σ , except for the duty factor and spreading losses (d_j/n_j) . For example, suppose the target has an RCS of 10 m^2 , the radar wavelength is 0.05 m , corresponding to a frequency of 6 GHz , and the jammer has a duty factor of 1.0 in a single detection cell ($d_j = n_j = 1$). The target gain in this case is approximately 47 dB ; if the jammer net repeater gain also equals 47 dB , it will match the target return. If the jammer must achieve a 10-dB JSR relative to that target, it will need a net gain of 57 dB , and so on. The net repeater gain can be achieved through any realizable combination of antenna gain and internal electronic gain.

12.3.6 Coherent Jammer Computation Example

To illustrate the application of the coherent jammer formulas, consider an example, using the same hypothetical, 6-GHz ground based radar and 10-m^2 target as in the noise jammer example of Section 12.3.4. For this case, we assume an SSJ (or co-located ESJ) with a linear repeater. The jammer has 3-dBi receive and transmit antennas (including polarization and other losses) and a small-signal electronic gain of 70 dB . The effective repeater throughput gain at the radar polarization is therefore $3 + 3 + 70 = 76 \text{ dB}$. The jammer has a peak transmitter power of 10 W , is repeating 100% of every pulse ($d_j = 1$), and is spreading its energy across 16 Doppler bins, corresponding to a 12-dB spreading loss $(1/n_j)$. The 10-m^2 target has an apparent target gain $(4\pi \sigma / \lambda^2)$ of 47 dB at 6 GHz . Therefore, in the gain-limited region, we expect the JSR to equal $76 - 12 - 47 = 17 \text{ dB}$ (throughput gain minus spreading loss minus target gain in dB). The 10-W transmitter power equates to 10 dBW . With a 70-dB small-signal gain, a received power level of -60 dBW ($= 10 \text{ dBW} - 70 \text{ dB}$) will cause the amplifier to reach its peak output power.

Figure 12-7 plots the signal, jammer, and thermal noise power. In the gain-limited region, the jammer power varies as R^{-4} just as the target does. As a result the JSR in this

FIGURE 12-7 ■
Coherent jamming
power example.



region is constant. We see that the JSR is indeed 17 dB as expected. The range at which the amplifier reaches its peak power is approximately 12.5 km. From that point forward, the jammer is in a power-limited region, and its power varies as R^{-2} , increasing more slowly than the target skin return as range decreases. Thus, the target signal eventually matches the jammer level at approximately 2 km and overtakes it at closer ranges (although that may be of questionable value in this case).

12.3.7 Jammer Received Power Computation

An important consideration in the design of transponders, repeaters, and ES systems is the detection sensitivity. The sensitivity is generally specified as the minimum detectable received power from the radar, expressed as either the receiver sensitivity or the system sensitivity. The receiver sensitivity is referenced to the power at the receive antenna port, i.e., the input to the RF front end of the EW system; this is typically expressed in dBm. The system sensitivity takes into account the receive antenna gain as well; this is often also expressed in dBm but is more correctly expressed as dBmi—the equivalent power that would be received through an isotropic (unity gain) antenna.

The formula for the constant gain repeater in Equation 12.12 implicitly includes the received power in the bracketed set of terms preceding $g_{j,el}$. The jammer receive power is thus

$$P_{j,rx} = (P_t G_{tj}) \left(\frac{\lambda}{4\pi R_j} \right)^2 G_{j,rx} \quad (12.17)$$

The received power computation from this formula is compared directly to the EW system sensitivity specification (because antenna gain is included) to determine whether the EW system has sufficient sensitivity to detect the radar under the assumed conditions. High radar ERP toward the jammer, high jammer antenna gain, and close-range operation increase the received signal and therefore improve the probability of detection by the EW system.

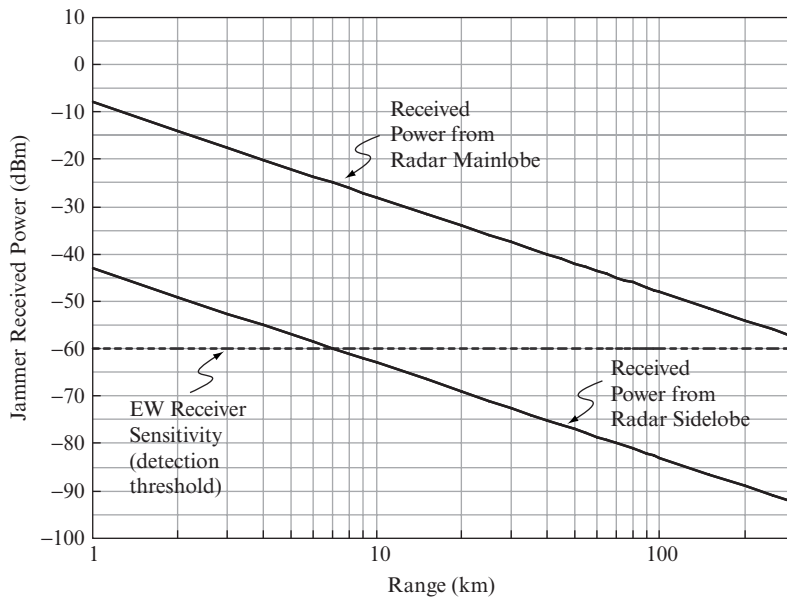


FIGURE 12-8 ■
Jammer received
power example.

12.3.8 Jammer Received Power Example

To illustrate the application of the jammer received power computation, continue the example from Table 12-2. We will consider only the ESJ (not the SOJ), and we will compare the power the ESJ receives in the radar main lobe with the power it receives in the radar sidelobe. Suppose the EW receiver has a sensitivity of -60 dBm, corresponding to the minimum detectable signal power present at the receive antenna port. Figure 12-8 plots the received power from the radar main lobe and sidelobe as a function of range and also marks the -60 dBm detection threshold in the EW receiver. The EW receiver is able to detect the radar beyond 200 km in the main lobe but not until approximately 7 km in the sidelobe. This corresponds to the 35-dB difference between the main lobe gain, 30 dBi, and sidelobe gain, assumed to be -5 dBi, as per Table 12-2. Thus, it would not be practical for this EW system to be employed in the radar sidelobes, unless very short-range SIJ roles are envisioned.

12.4 | EP OVERVIEW

Electronic protection is intended to reduce the effectiveness of an opponent's EA or ES capability. Many radar design features that are motivated by the need to improve performance under stressing target, clutter, and electromagnetic interference (EMI) conditions also provide inherent resistance to EA. Examples include low radar sidelobes, auxiliary antenna blanking and cancellation channels, monopulse angle measurements, pulse compression, wide dynamic range, Doppler processing, and frequency diversity. These design features are always in effect, and they provide optimum performance under both natural and EW environments. Such design features can be described as providing intrinsic EP. In contrast, some EP features are only enabled under suspected EW conditions. These features can be described as responsive EP. Responsive EP does not in general provide

optimal performance under natural conditions, and is consequently used only when needed. In addition, responsive EP may require some form of detection, characterization, and assessment of the EW environment to determine whether to initiate or terminate the selected EP. The inherent time lag associated with the environment sensing and EP decision logic may lead to improper EP implementation in a dynamically changing EW environment. Radar performance may actually suffer more due to the sluggish EP switching than it would without any EP.

Numerous EP techniques and designs have been developed over the decades. The ones presented in this chapter provide at best a small sampling of the wide variety of approaches that have been conceived and implemented. The radar system engineer's selection of an appropriate suite of EP techniques must take into consideration a wide variety of factors, such as EW environment severity and likelihood; EP technology maturity; EP design impacts on radar performance under natural environments; radar size, weight, cooling, and power limitations; and program schedule and funding constraints.

The reactive nature of EW makes it critical to continually understand and anticipate an adversary's capabilities in order to successfully counter them. The terms robust and durable are sometimes used to describe an EP capability that provides consistently good performance over a wide range of EA conditions and cannot be quickly defeated by a redesigned or reprogrammed EW threat once its existence and characteristics are revealed. In contrast, the terms fragile and perishable are sometimes used to describe an EP capability whose performance breaks down when specific threat assumptions are not met or that can readily be countered once it has been exposed.

Electronic warfare concerns are at times given secondary priority in radar development efforts. This might be attributable to a lack of understanding of the wide range of potential EW environments and the ease with which reactive EW threats may emerge. In some cases, addressing EP requirements might be deferred until after a baseline radar development is under way or even completed to avoid compromising a program's survival through the revelation of potential performance shortfalls. Generally, however, modern radar development identifies and incorporates critical EP requirements from the earliest stages of the design to best ensure that the final system is properly equipped to perform its mission under any anticipated electromagnetic environment.

The EP techniques described in the subsequent sections are organized into categories corresponding to major radar functional subsystems: antenna, receiver, transmitter, exciter, signal processor, and data processor. Many of the EP techniques obviously have design implications across multiple subsystems; in such cases we pick the subsystem that seems to be most central to conveying the EP concept. The actual division between the categories themselves may be somewhat blurred in modern radars. For example, radars may incorporate parts of the transmitter and receiver in the antenna array or may digitally implement traditional analog receiver functions in the signal processor. Therefore, the categories are intended primarily in a functional sense rather than an actual physical implementation.

12.5 | ANTENNA-BASED EP

The antenna is the radar's first line of defense against jamming. Electronic protection is accomplished through pattern attenuation and blanking of EA interference as well as angle tracking approaches that help counter angle deception EA.

12.5.1 Low Sidelobes

Low antenna sidelobes provide effective EP against both noise and deceptive sidelobe interference. Low sidelobes in the receive antenna pattern force sidelobe jammers to have higher power to achieve a desired JNR or JSR. This is evident through inspection of Equations 12.9, 12.13, and 12.14 for the noise and coherent jammers, where a reduction in G_{rj} corresponds directly to a reduction in JSR. Low sidelobes in the transmit antenna pattern, G_{tj} , force repeater and transponder jammers as well as receive-only ES systems to have increased sensitivity to detect and characterize the radar transmission, as described in Section 12.3.7.

One of the limiting factors in repeater and transponder operation is the T/R isolation. The jammer T/R isolation requirement is directly impacted by both transmit and receive radar sidelobes. For example, if the radar reduces its transmit sidelobes by 5 dB and receive sidelobes by 10 dB, the jammer T/R isolation requirement increases by 15 dB.

Theoretically it is possible to achieve very low sidelobe levels through appropriate weighting functions across the radar aperture. In practice, however, limitations in design, fabrication, calibration, and component stability make realization of such levels elusive. Solid-state phased array antennas, for example, have multiple potential contributors to the antenna sidelobe level, including amplitude and phase errors, T/R module attrition, and correlated errors between groups of elements in the array. Reducing antenna sidelobes must be weighed against any accompanying reduction in main lobe directivity and widening of the main beam. Often it is preferable to employ a uniform aperture weighting on transmit to achieve the maximum ERP while relying on the receive patterns to provide the sidelobe reduction benefits.

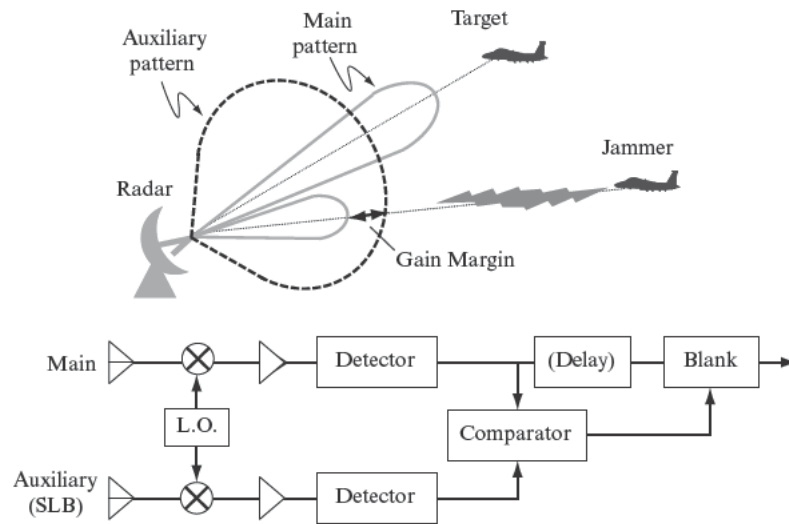
12.5.2 Sector Blanking

Sector blanking offers some EP benefit to search radars that are subjected to EA interference from known, specific angular regions. The radar avoids receiving or both transmitting and receiving in the defined angular sector. Although such a mode is tantamount to the radar conceding defeat in the blanked sectors, there may be some benefit as well. One benefit is that the radar might be able to prevent a sidelobe transponder or repeater from detecting and locking on to the radar frequency or synchronizing to the antenna scan pattern by depriving the jammer of a main lobe illumination [3,10]. Sector blanking might be part of the operational tactics, techniques, and procedures (TTPs) that are developed for the radar.

12.5.3 Sidelobe Blanking

Sidelobe blanking helps eliminate sidelobe false targets that are strong enough to overcome the antenna sidelobe pattern attenuation [6,10]. The concept is depicted in Figure 12-9. The radar employs one or more broad-beam, auxiliary antennas in addition to the main antenna. The auxiliary antenna has higher gain than the main antenna in the region of the main antenna sidelobes and lower gain than the main antenna in the region of the main antenna main lobe. The auxiliary antenna is connected to an independent radar receiver channel that is essentially identical to that used for the main antenna. A comparison is made of the main and auxiliary channel detected signal amplitudes. If the auxiliary channel signal exceeds the main channel signal by some margin, the signal in the main channel is blanked;

FIGURE 12-9 ■
Sidelobe blanker
concept.



if the signal in the main channel exceeds that of the auxiliary channel by some margin, the signal in the main channel is processed. A non-EW analog to the sidelobe blanker (SLB) is the guard channel used by airborne pulse-Doppler radar to reject strong discrete clutter returns in the sidelobes during medium pulse repetition frequency (PRF) operation [6,9].

There are several design challenges associated with practical implementations of an SLB. First, it may be difficult to design an auxiliary antenna whose pattern is broad enough to cover the entire sidelobe region of the main antenna while also having sufficient gain to cover the main antenna's near-in sidelobes and various error-related sidelobes. Sometimes multiple auxiliary antennas or steered SLB arrays are used to address this problem. Second, because thermal noise is independent in each detection channel, a noise-induced false alarm in the SLB channel may cause blanking of a target whose signal is reduced by destructive interference from thermal noise in the main channel, thereby reducing probability of detection. This underscores why the SLB antenna gain cannot be enhanced with additional electronic amplification: the result would be an increase in this so-called blanking loss. Third, in the absence of a target signal, an instance of high thermal noise in the main channel combined with thermal noise-induced destructive interference of an SLFT in the blanker channel may cause the false target not to be blanked (main exceeds SLB, but no target exists). Finally, the SLB becomes counterproductive under conditions of high duty cycle interference: the radar probability of detection would be severely compromised by excessive blanking unless either a limit were placed on the maximum blanking rate or the SLB were integrated with a sidelobe canceller function.

12.5.4 Sidelobe Cancellation

Sidelobe cancellation (SLC) is a means of reducing continuous or high duty cycle noise interference from a sidelobe jammer [1,6,9,10]. One or more auxiliary sidelobe canceller antennas are used, each with a dedicated receiver channel. The auxiliary antennas generally have a broad pattern, similar to that of a SLB antenna. Sidelobe noise jamming that is strong enough to obscure a main lobe target will produce a much stronger signal level in the auxiliary channels than the target return. The gain and phase of the auxiliary channels

are adjusted with complex (amplitude and phase) weights adaptively derived from the interference signals being received such that the jamming interference received through the main antenna is minimized. This effectively generates spatial nulls in the directions of the interference, with the intent of simultaneously preserving the underlying main lobe target signal. The cancellation ratio, which is the ratio of the uncanceled signal to the canceled signal, may be tens of dB and depends on factors such as the waveform bandwidth and the accuracy with which channels are gain and phase matched. Another measure of performance is the change in signal-to-interference-plus-noise ratio (SINR), which reflects both the reduction in jammer interference as well as possible unintended loss in desired signal. In general there must be at least one auxiliary channel for each spatially separated interference source in the environment, although multipath scattering may require additional channels.

Early, analog implementations of SLC featured closed-loop feedback control circuits that adjusted the weights of the auxiliary channels within a certain time period. The SLC weights therefore had an associated settling time that introduced a lag to changes in the environment. Many modern radars employ a digitally based approach whereby the gain and phase adjustments are applied numerically in the signal processor and are determined independently on each data batch being processed. Adaptive digital beamforming (DBF) is a design approach in which an electronically scanned array (ESA) antenna is divided into subarrays, with each subarray feeding an independent receiver channel whose output is digitized through an ADC. The subarrays may include hundreds or even thousands of radiating elements; or, in the limit, there may be only one element per channel. The digitized channels can be combined in multiple ways on the same sample of data to extract the desired information without consuming additional radar resource time. Refer to Chapters 9 and 10 for a more complete discussion of adaptive cancellation of sidelobe interference.

12.5.5 Main Lobe Cancellation

Main lobe cancellation (MLC) is used to adaptively cancel one or more escort noise jammers in the main beam. The MLC places an adaptive spatial null in the direction of the escort jammer(s) while trying to maintain acceptable pattern gain elsewhere in the main lobe region to allow detection of the nearby target. Although the pattern gain toward the target may be reduced in the process, the resulting SINR still improves due to the even greater attenuation of the jammer(s). The SINR improvement resulting from MLC decreases as the jammer-to-target angular separation decreases; for separations on the order of 0.1 beamwidth, it may be difficult to cancel the jammer without also significantly attenuating the target. Typically MLC apertures must be much larger than SLC auxiliary antennas because the target signal level is much stronger in the direction of main lobe interference than for sidelobe interference.

The process of generating a null in the main beam may cause the average sidelobe levels to increase. This might be a concern if strong sidelobe clutter were present or if there were other jamming or EMI in the sidelobes. Another potential consequence of MLC is the degradation in the angle estimate of the underlying target. For example, an angle estimate formed from a monopulse antenna is based on the designed and calibrated relationship between the complex monopulse ratio of the difference (Δ) and sum (Σ) beams as a function of the angle off boresight, as described in Section 12.5.9. If the Δ and Σ patterns are each affected differently by the MLC process, their ratio will in turn also

be different, thereby resulting in an erroneous angle measurement. Additional calibration and processing algorithms may be needed to enhance MLC if such errors are unacceptably high. Refer to Chapters 9 and 10 for a more complete discussion of adaptive cancellation of main lobe interference.

12.5.6 Narrow Beamwidth

A narrow antenna beamwidth reduces the angular region of influence of a main lobe support jammer. An ESJ will have to fly closer to the target(s) being protected to remain in the main beam; this may impact tactics or require additional ESJs to screen multiple targets. The angular region in which an SOJ or SIJ is in the main beam simultaneously with the target is also reduced. The narrow beamwidth also implies high transmit and receive antenna directivity. The transmit directivity increases the radar ERP which decreases the JSR as shown in Equations 12.9, 12.13, and 12.14. The receive directivity increases the main lobe to sidelobe ratio, thereby making it more difficult for a sidelobe jammer to penetrate the radar sidelobes. Such EP benefits notwithstanding, the selection of beamwidth in a radar system design is rarely if ever motivated by EW considerations but rather by search volume coverage, sensitivity, and angle estimation accuracy requirements.

12.5.7 Random Beam Scanning

Random beam scanning is a capability afforded by ESAs that can reduce the effectiveness of transponder and repeater jammers attempting to operate in the radar sidelobes. Electronically scanned array radars have the ability to generate arbitrary scan patterns when performing search. In contrast, mechanically scanned antennas generally follow predictable raster or circular scan patterns with regular, periodic revisits to a specific beam position. A jammer operating in the sidelobe can potentially exploit such predictability and generate its interference, whether noise or false targets, at specific angles by synchronizing to the periodic radar main beam illumination. For example, a mechanically scanned antenna with a 0.2-Hz circular scan pattern will paint the jammer with a main lobe illumination once every 5 seconds. If the jammer knows that the radar uses a clockwise scan, it can place its SLFTs at, say, 36 degrees further clockwise from its own position by timing its response to occur 0.5 seconds after receiving the main lobe illumination ($36/360 \times 5.0 \text{ sec} = 0.5 \text{ sec}$). The ESA randomization of the scan pattern denies the jammer this capability, forcing it to employ a potentially less effective EA mode.

12.5.8 Passive Conical Scanning

Prior to the arrival of monopulse antennas, radars often employed conical scan (CONSCAN) antennas to perform angle tracking. In a traditional, active CONSCAN approach, the antenna beam is nutated about an axis that is intended to be at or near the direction of the target being tracked. As the center axis drifts away from the target position, the amplitude of the target return is stronger for beams closest to the target and weaker for beams farther from the target. This produces an AM on the received target signal. The phase of this AM signal relative to the CONSCAN modulation control voltage indicates the direction of the angle tracking error, and the relative amplitude variation indicates the magnitude of the correction required. The angle error estimate can only be obtained by taking measurements from multiple pulses over the modulation scan period.

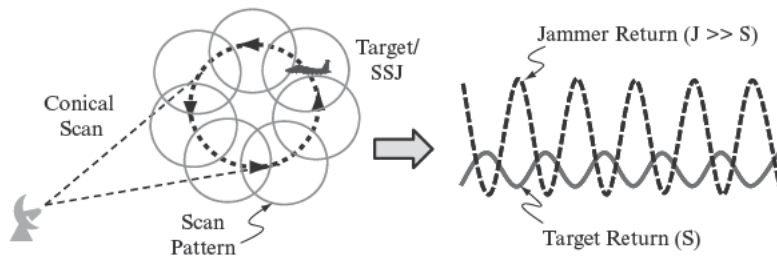


FIGURE 12-10 ■
Inverse gain angle
deception against
conical scan radar.

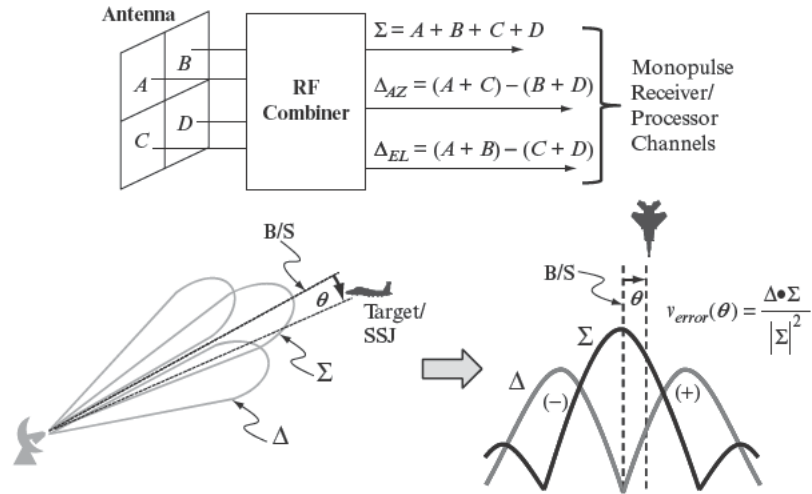
A self-protection jammer uses inverse gain (IG) to exploit the active CONSCAN process, as illustrated in Figure 12-10 [2,10]. The jammer reduces its transmit power when it receives a strong signal from the radar, indicating the radar beam is close to the jammer, and the jammer increases its transmit power when it receives a weak signal from the radar, indicating that the radar beam is far from the jammer. The result is that the radar receives an angle track error signal from the jammer that is opposite to what would normally occur from a target at the same position, causing the antenna to steer further and further away from the jammer. The jammer can vary the phase of its AM cycle relative to the received radar scan modulation to gradually walk the angle off if desired.

Two variations of CONSCAN provide EP against IG but still leave the radar vulnerable to AM-based jamming. Scan-on-receive-only (SORO), also referred to as conical-scan-on-receive-only (COSRO), and lobe-on-receive-only (LORO) employ passive scanning rather than active scanning; that is, they extract angle information by either nutating or stepping the receive beam (passive) but do not scan the transmit beam (active) [2,3,10]. Jammers therefore cannot use the received signal power from the radar to determine how to generate a synchronized AM for angle deception. As a result, the jammers resort to a swept AM (SAM) approach, whereby AM is generated at a frequency that is varied over a range that encompasses the expected scan rate of the radar. The AM rate might be swept from 1 Hz to 5 Hz, for example. When the jammer AM rate happens to be close to or at the radar scan rate, angle errors might be induced. This might be combined with a jog detector which senses a drop in received power from the radar indicating the transmit beam being steered away from the target and therefore an effective AM rate being found.

12.5.9 Monopulse Angle Measurement

The development of monopulse antennas proved to be a major advance for radar EP against angle deception countermeasures. Specifically, monopulse completely neutralized the AM-based techniques such as IG and SAM that targeted active or passive conical scanning radar. The monopulse angle measurement method, described in [8], does not rely on a multiple-pulse scan modulation but instead forms a ratio of the signals received through the difference, Δ , and sum, Σ , patterns, thereby supplying the angle information on each received pulse. A monopulse antenna is divided into four quadrants, A, B, C, and D, as shown in Figure 12-11. The quadrants are fed into an RF combiner that produces the sum ($A + B + C + D$), azimuth difference ($A + C - B - D$), and elevation difference ($A + B - C - D$) outputs. These outputs are fed into independent receiver/processor channels which form the complex ratio of the difference-to-sum voltages to provide the basis of the angle error signal, $v_{error}(\theta)$ for a given angle off boresight (B/S) of θ . The complex voltage is the ratio of magnitudes multiplied by the cosine of their relative phase, β , or, equivalently, the dot-product of the difference and sum voltages normalized by the

FIGURE 12-11 ■
Monopulse angle
measurement
concept.



sum magnitude squared:

$$v_{error}(\theta) = \frac{|\Delta|}{|\Sigma|} \cos \beta = \frac{(\Delta \cdot \Sigma)}{|\Sigma|^2} = \frac{(\Delta \cdot \Sigma)}{\Sigma \cdot \Sigma} \quad (12.18)$$

Although some radar modes such as moving target indication (MTI) and pulse-Doppler require integration of multiple pulses before extracting angle measurements, the monopulse architecture still imparts the complete angle information on each received pulse.

The monopulse-based angle estimate is immune from AM jamming because the Δ/Σ ratio is independent of signal amplitude: any AM generated by the jammer appears in both the difference voltage (numerator) and sum voltage (denominator) and is therefore cancelled out in the ratio. A hybrid of conical scan and monopulse called conopulse, or scan with compensation (SWC), has similar immunity to AM [3]. The conopulse antenna uses two simultaneous lobes that are nutated around a central axis, each connecting to a receiver channel. The ratio of the channels removes any jammer or target AM. This is essentially a two-channel version of monopulse.

12.5.10 Low Cross Polarization Antenna

A low cross polarization antenna helps reduce the susceptibility of monopulse radar to XPOL angle deception. The monopulse sum and difference patterns are quite different in the orthogonal polarization than they are in the principle polarization. The orthogonal sum pattern, for example, may have a null on boresight instead of a peak. This would certainly impact the monopulse angle estimate that is formed by taking the Δ/Σ ratio. Returns from the environment that are co-polarized to the radar generate difference and sum voltages that correspond to the intended monopulse patterns, while returns from the environment that are cross-polarized to the radar generate voltages that correspond to distorted patterns. The angle error signal used by the radar therefore contains two conflicting components: one providing a correct error signal and the other providing an erroneous error signal. Fortunately for the radar, the cross-polarized pattern gain is much lower than the co-polarized gain and therefore does not have a significant effect for most returns of interest. Thus, natural target returns do not cause significant

polarization-induced angle errors even though they may contain cross-polarized components that are comparable to their co-polarized components.

An XPOL jammer generates a signal that is approximately orthogonal to the receive polarization of the radar. The jammer must have very high JSR to compete with the co-polarized target skin return after being attenuated by the polarization mismatch. The XPOL jammer must, in a sense, also compete with itself as small departures from the desired polarization contain co-polarized components that act as a beacon for the radar. A low cross-polarization antenna acts to reduce the cross-polarized antenna pattern levels relative to the co-polarized patterns. This forces the jammer to increase both its transmit power (for higher JSR) and its polarization control accuracy (less co-polarized contamination), both of which may significantly stress jammer design. Parabolic dish antennas are notorious for having high cross-polarization patterns. Migrating away from such antennas to a flat-plate array, for example, may help reduce the effects of XPOL jamming [1].

12.6 TRANSMITTER-BASED EP

The transmitter determines the radar's peak and average radiated power. The choice of transmitter is closely tied to the choice of radar waveform in determining duty cycle, maximum pulse width, IBW, and tunable bandwidth. Potential EP benefits can be obtained through either increasing or decreasing the radiated power of the radar.

12.6.1 High Peak Power

High transmitter peak power directly reduces the achievable JSR of a noise jammer or constant-power repeater. This is evident by P_t in the denominator of Equations 12.9 and 12.13 for those jammer types. Some angle deception EA techniques such as IG, AGC deception, XEYE, XPOL, and TB require very high JSR to overcome pattern loss, duty factor loss, spatial nulling loss, polarization loss, and bounce loss, respectively. High peak transmit power therefore affects the feasibility and affordability of these EA techniques. As long as the jammer is able to operate in a constant-gain mode, however, increasing the transmitter power does not provide an EP benefit, except for the fact that it increases the range at which such a jammer will become peak-power limited.

In practice, the selection of the radar transmitter has little or nothing to do with EP considerations, and therefore any EP benefits are considered fortuitous. The radar transmitter is one of most costly components in the radar system and is generally selected based on available technology and target detection requirements. Factors such as cost, weight, volume, cooling, and power consumption sometimes place severe constraints on transmitter selection, especially in airborne applications. Even if flexibility were provided to increase transmitter power, there are two accompanying disadvantages from doing so from an EW perspective: increased detectability by ES receivers and increased vulnerability to ARMs.

12.6.2 Transmit Power Reduction

Reducing the peak transmit power to the minimum necessary for target detection may help prevent EW transponders and repeaters from detecting and responding to the radar emissions. This is more easily achieved when the radar engages high-RCS targets at close range, for example, where there is SNR to burn. The benefits are especially pronounced

against sidelobe EW systems as these systems may already be stressed in detecting radar sidelobe emissions. We can estimate the minimum transmit power by rewriting the radar range equation, Equation 12.5, as

$$P_{t,min} = \frac{SNR_{min} L_s (4\pi)^3 R^4 kTF}{G_t G_r \lambda^2 \sigma n \tau} \quad (12.19)$$

where $P_{t,min}$ is the minimum transmit power necessary to detect a target of a given RCS, σ , and range, R , and SNR_{min} is the minimum SNR necessary for detection. The resulting value of $P_{t,min}$ can then be used in the expression for the jammer received power shown in Equation 12.17 to see if the signal received by the jammer exceeds its detection threshold.

12.6.3 Emission Control

Emission control (EMCON) is a more dramatic way of reducing radar emissions than peak power reduction. It involves restricting radar transmissions to only the minimum amount of time necessary as well as to only the angular sectors necessary. This helps prevent EW systems from detecting, identifying, and locating the radar [3,5,10]. It also reduces the radar vulnerability to ARMs that may be unable to accurately coast for long intervals without an angle measurement for guidance. A mobile radar might use EMCON combined with rapid relocation to further defend against weapon attacks.

Emission control using short, infrequent transmissions might be able to delay the detection by EW systems or ARMs that employ scanning antennas or scanning narrowband receivers. Such systems may be sacrificing probability of intercept (POI) in favor of improved sensitivity. Probability of intercept refers to the probability that the radar emission is within both the antenna instantaneous field of view and receiver IBW of the EW system. Narrow beamwidth (higher spatial gain) and narrow IBW (less thermal noise) increase sensitivity but decrease POI, whereas wide beamwidth and wide IBW increase POI but decrease sensitivity. Channelization of the EW receiver helps simultaneously preserve both POI and sensitivity in the receiver but at the expense of higher cost, weight, size, and complexity.

12.7 | EXCITER-BASED EP

The radar exciter determines the characteristics and diversity of the radar waveforms that are generated. Potential EP benefits are realized through long-duration waveforms, range-resolution waveforms, and waveform diversity and agility.

12.7.1 Wide-Pulse Waveform

Wide-pulse, or long-pulse, waveforms provide an alternative to high peak transmit power as a means of increasing the total energy of a target relative to noise jamming. The JSR of a noise jammer is inversely proportional to the pulse width, as seen in Equation 12.9. Some transmitters have a maximum pulse width capability, and therefore a wide-pulse waveform must be compatible with the transmitter design. Range resolution is also a possible concern. The range resolution, ΔR , for a simple, unmodulated pulse is equal to $c\tau/2$. Going from a 1- μs pulse to a 10- μs pulse, for example, would reduce the JSR by 10 dB but would also decrease the range resolution from 150 m to 1500 m,

potentially an unacceptable compromise. Pulse compression, discussed in Section 12.7.5, allows the radar to use a wide pulse while simultaneously maintaining the required range resolution.

12.7.2 Multiple-Pulse Waveform

Another method of increasing the total duration of the radar waveform in addition to the use of a wide pulse is through multiple-pulse waveforms. Multiple-pulse, coherent waveforms such as MTI and pulse-Doppler provide inherent resistance to noise jamming because they increase the total energy of the signal relative to noncoherent noise. The JSR of a noise jammer decreases relative to that of a single pulse by a factor approximately equal to the number of pulses coherently integrated, n , as seen in Equation 12.9. (This is approximate because additional processing losses, such as amplitude weighting across the pulses, may be incurred in integrating the multiple pulses.)

In general, the choice of specific modes such as MTI and pulse-Doppler is dictated by the radar performance objectives and the clutter environment rather than EP considerations. In fact, there are actually some potential disadvantages of these waveforms from an EP perspective compared with single-pulse waveforms. The multiple-pulse dwells at a constant frequency provide ample time for a responsive noise jammer to detect, acquire and set-on to the radar frequency, even for some SOJ (long-range jammer screening a shorter-range target) geometries. In addition, long repetitive pulse-Doppler dwells at a constant PRI allow coherent jammers to perform PRI tracking and generate apparent up-range false targets by exploiting the range ambiguity of such waveforms.

12.7.3 Burnthrough Waveforms

Burnthrough waveforms refer to waveforms that employ wider pulse widths or a greater number of pulses than are used in the normal radar waveform set, explicitly for the purpose of increasing the target energy in the presence of noise jamming [2,5,10]. (The term burnthrough also denotes the effect of the target skin return emerging out of the jamming noise as the radar-to-target range decreases; this is not an EP technique per se, however.) Burnthrough may be achieved by increasing the radar average power, increasing the total time on target, or both. For a given PRF, f_p , increasing τ and n is equivalent to increasing the duty factor, d , and coherent processing interval (CPI) time, T_{cpi} :

$$\tau n = (\tau f_p) (n T_p) = d T_{cpi} \quad (12.20)$$

where T_p is the radar PRI ($T_p = f_p^{-1}$). The CPI time is the total time of the waveform—whether single or multiple pulse—that is coherently processed by the radar. (Some radar modes transmit more pulses than are actually processed in a CPI, due to clutter-fill requirements, for example.) Post-detection (noncoherent) integration may be used instead of, or in addition to, coherent integration as part of a burnthrough waveform.

Burnthrough may place a high burden on the resource allocation of a multifunction ESA: a 3-dB burnthrough improvement requires twice the normal time on target, a 6-dB improvement requires quadrupling the time on target, and so on. Clearly, burnthrough cannot be used excessively without compromising the radar's ability to perform its scheduled tasks in a timely manner.

As is the case for conventional multiple-pulse waveforms, burnthrough waveforms have the disadvantage of providing responsive noise jammers and coherent repeaters

additional time to set-on to the radar waveform and generate more efficient jamming. Two additional factors that may limit the duration of a burnthrough waveform are range-walk and Doppler-walk. Range-walk occurs when a high range-rate target moves through multiple range bins over the duration of the burnthrough dwell. This is more likely for a wideband waveform used for high range resolution. Similarly, Doppler-walk occurs when a high-acceleration target moves through multiple Doppler filters over the course of the dwell. The signal processor may be able to employ algorithms to compensate for the range-walk and Doppler-walk issues.

12.7.4 Narrow-Pulse Range Resolution

Range resolution accomplished through a narrow pulse, or short pulse, can help counter transponder or repeater SSJs that have relatively slow response times or long throughput delays [6]. A responsive noise SSJ may rely on detection of the pulse leading edge to synchronize its transmission with the radar waveform. It may also attempt to sample the radar frequency within the early portion of the pulse in order to set-on with RSN. If the pulse is narrow enough, say less than 300 ns, the jammer may not be able to place its signal on top of the skin return of the target due to these delays. This is especially true for a radar with PRI agility and pulse-to-pulse frequency agility, in which case the noise jammer must dilute its energy with broadband noise jamming and may have to broadcast continuously rather than in a cover-pulse mode.

A coherent jammer may also have difficulty with very narrow pulses. The cumulative insertion delay from transmission lines to and from the antennas as well as various components in the repeater chain may result in a total throughput delay of 200 ns or more. Against pulse widths of 1 μ s, for example, such delays might be tolerated in the jammer design, as most of the radar pulse is still covered. Against narrow pulses, however, the jamming may cover only a small fraction of the radar pulse or may miss the pulse completely. Other examples of EA techniques with time-delayed returns are TB and TRD [6,9]. Jammers with coherent RF memory and PRI tracking may be able to generate a coherent cover pulse to compensate for this throughput delay if the radar has a predictable PRI waveform. Some repeater jammers have a minimum pulse constraint that may result from limitations in pulse parameter measurement, encoding, and processing functions, in which case the PRI tracking may never be accomplished.

The main disadvantage of the narrow-pulse mode is the potential sacrifice in average power against noise jammers if it is used instead of longer-pulse waveforms that are otherwise available to the radar. The target SNR is inversely proportional to the pulse width, as indicated in Equation 12.5. The narrow pulse does not compromise performance if it can be combined with a higher PRF to maintain average power or if a reduced average power can be tolerated due to high-RCS or short-range target conditions. The potential loss of SNR resulting from a narrow pulse is circumvented through the use of pulse compression waveforms, described below.

12.7.5 Pulse Compression Range Resolution

Pulse compression allows the radar to achieve a desired range resolution without sacrificing SNR [6]. As described in [8], instead of transmitting a narrow pulse that corresponds to the desired range resolution, the radar transmits a much wider pulse. The wide pulse contains an intrapulse PM or FM that results in a signal of the same spectral bandwidth as the

erstwhile narrow pulse. The increase in total energy (power multiplied by time) on the target allows the radar to increase its SNR as well as reduce the JSR of noise jamming. Pulse compression can be implemented through a variety of methods, most common of which are linear FM (LFM) and pseudo-random binary-phase coding. For both a narrow-pulse and a pulse compression waveform, the range resolution, ΔR , is inversely proportional to the waveform bandwidth, B , through $\Delta R = c/2B$, with the corresponding time resolution being equal to the inverse of the bandwidth. As for a narrow pulse already described, the range resolution achieved through pulse compression allows the radar to separate the skin return of the jammer platform from the leading edge of the jammer return if the jammer has a delayed response.

Aside from the preservation of SNR, a pulse compression waveform has some other potential advantages relative to a narrow pulse of comparable range resolution when operating against certain types of jammers. Coherent repeaters using recirculating RF memory loops have limited capability against pulse compression waveforms. Such jammers capture a fraction of the radar pulse in a fixed coherent delay line and continuously recirculate the captured signal multiple times within a memory loop. A coupled output of the memory loop is gated on and off to provide a variable time-delayed sample of the captured signal. This approach is reasonably effective against unmodulated pulses but will not preserve the intrapulse modulation of a pulse compression waveform. Pulse compression also forces the jammer to split its power if it spaces false targets more closely than the uncompressed pulse width. For example, if an SSJ wants to generate four RFTs beginning 300 m behind the target and occurring at 300-m spacing, the targets would be generated at delays of 2, 4, 6, and 8 μs relative to the target return. If the radar employs a 20- μs pulse compression waveform, all four RFTs will occur simultaneously for part of the time, thereby causing a dilution of transmitter power by nominally 6 dB ($20 \log 4$). Neither of these problems exist for the jammer operating against a conventional, unmodulated pulse.

Wideband pulse compression (and narrow-pulse) waveforms may also help prevent detection by a transponder or repeater. The jammer must have a receiver detection bandwidth wide enough to capture the entire radar waveform. Increasing the instantaneous bandwidth increases both the thermal noise and the density of other signals in the environment that must be filtered and sorted by the EW receiver/processor. Although typical EW receivers employ wide IBWs to handle multiple emitters as well as frequency agile radars, a wideband waveform may prevent use of a narrowband, high-sensitivity adjunct receiver.

One potential disadvantage of pulse compression relative to a narrow pulse is that a noise cover pulse jammer can no longer be resolved from the skin return. The NCP jammer has ample time to be triggered and set-on to the radar frequency, if applicable, within the initial portion of the uncompressed pulse width. The delayed jammer will not be resolvable from the compressed target return because the radar must process the entire pulse in order to achieve pulse compression and the jammer is present during a high percentage of the uncompressed pulse.

12.7.6 Doppler Resolution Waveform

In addition to their ability to reduce the JSR against noise jamming, long, multiple-pulse coherent waveforms enable the radar to resolve targets from clutter and EA returns. The Doppler resolution, Δf_D , is approximately equal to the inverse of the total CPI time, T_{cpi} .

The resolution in radial velocity, Δv , is directly proportional to resolution in Doppler:

$$\Delta v = \frac{\lambda \Delta f_D}{2} = \frac{\lambda}{2T_{cpi}} \quad (12.21)$$

For example, a 10-ms coherent pulse burst provides a Doppler resolution of approximately 100 Hz. At X-band, this corresponds to a difference in radial velocity Δv of approximately 1.5 m/s. One benefit of Doppler resolution is the ability to resolve a chaff cloud from an airborne target based on the velocity difference. The chaff slows down quickly after being released and can easily be separated in the Doppler domain. The same benefit may apply to unpowered expendable decoys.

12.7.7 Frequency Variation

Frequency variation is one of the most effective EP features for countering BN jamming. The benefit arises from forcing the noise jammer to spread its energy over a wide bandwidth, thereby decreasing the power spectral density at the radar's instantaneous operating frequency. This is evident in Equation 12.9, where the JSR for a noise jammer is seen to decrease as the jammer bandwidth, B_j , increases.

Frequency variation can be achieved through either frequency diversity or frequency agility. Frequency diversity refers to the ability of the radar operator to switch operating frequencies over a given RF range. For example, the radar operator may be directed to select one of several possible channels or oscillators prior to or during a mission for reasons such as avoidance of EMI or improving electromagnetic compatibility (EMC) with other systems. Frequency agility refers to the real-time, automatic variation of frequency on a pulse-to-pulse, burst-to-burst, or scan-to-scan type of time frame. In addition to forcing a barrage jammer to spread its energy, frequency agility makes it more difficult for a repeater or transponder to keep the radar within its temporal and spectral tracking windows. Frequency agility can be combined with an analysis of the jamming spectrum and a selection of the least-jammed frequency, as described in Section 12.8.8.

Pulse-to-pulse frequency agility has the greatest EP benefit because it limits a support jammer's ability to screen targets at ranges closer than its own position without resorting to BN. This greatly reduces the effectiveness of an SOJ because it cannot concentrate its noise power density through RSN type waveforms. An SSJ or ESJ that attempts to employ an RSN technique may also announce its own range through the leading edge of its noise return. Unfortunately, pulse-to-pulse frequency agility is not generally available to radars operating in high clutter environments, as MTI and pulse-Doppler waveforms traditionally employ constant-frequency pulse-burst waveforms.

12.7.8 Multiple Simultaneous Frequency Radiation

A radar can potentially counter transponders and repeaters by simultaneously transmitting at multiple frequencies. This could force the jammer to dilute its power over a much wider bandwidth than the normal radar bandwidth or may cause a jammer to miss one or more of the radar frequencies. For example, suppose the radar normally uses a 1-MHz bandwidth waveform, but under the multiple simultaneous frequency radiation mode it radiates 10 such waveforms of comparable amplitude, spaced 100 MHz apart. One possible outcome is that a noise transponder reverts to a BN waveform instead of RSN waveform. This would cause it to spread its energy over a 1000-MHz BN bandwidth instead of perhaps a 5-MHz

RSN bandwidth. The jammer power is thus diluted by 23 dB (a factor of $1000/5 = 200$), whereas the radar power per frequency is diluted by only 10 dB, corresponding to the 10 distinct frequencies: a 13 dB reduction of JSR. A more efficient jammer may be able to generate multiple spot noise waveforms, however, thereby incurring only a 10-dB dilution loss and resulting in no net advantage to the radar.

Another possible result is that the jammer lacks sufficient bandwidth to detect and respond to all frequencies simultaneously and may therefore miss one or more frequencies entirely. If the radar is able to receive and process these frequencies, it may be able to get an unjammed look at the target. A jammer may overcome this by time-sharing between the multiple frequencies if the radar uses multiple simultaneous frequency radiation in conjunction with a multiple-pulse waveform.

The multiple simultaneous frequency radiation capability is a potentially costly feature. For sidelobe operation, it may involve auxiliary transmitters and antennas. For main lobe operation, it may involve multiple exciters and transmitters or it may force a shared transmitter to operate in its linear region in order to avoid intermodulation products, thereby reducing radiated power.

12.7.9 Pulse Repetition Frequency Variation

The variability of the radar PRF drives up the complexity of a transponder or repeater that is required to identify and possibly temporally track radars in the environment. The radar PRF can be varied pulse to pulse or, in the case of a pulse-Doppler radar, burst to burst (sometimes referred to as dwell to dwell). Pulse-to-pulse jitter introduces small, random-like variations in the radar PRI; pulse-to-pulse stagger introduces regular switching between a several specific PRIs. Deceptive jammers attempting to project false targets in front of their true range will be unable to do so if the radar PRI is not predictable from pulse to pulse. For burst-to-burst variations, the jammer must first measure the new PRI and adapt its EA transmission accordingly. It may also have to correlate the previous and current PRIs to the same emitter being countered before resuming the EA technique. These effects may potentially interrupt and delay the jammer response with each PRF change.

12.8 | RECEIVER-BASED EP

The receiver encompasses the signal path that accepts the RF input from the antenna, translates the signal to one or more IF stages, and finally converts the signal to digital words through an ADC. Many of the receiver design practices used to optimize radar performance under natural environments, such as those related to dynamic range and spectral filtering, also provide inherent EP benefits. Receiver EP capability may be further enhanced to handle specific types of jamming through specialized gain and filtering methods or additional detection channels. A description of radar receiver design is provided in [8].

12.8.1 RF Preselection

The RF preselector is a front-end RF filter that passes the frequency spectrum of the radar waveform but rejects out-of-band interference. The radar receiver–signal processor chain typically employs band-limiting IF filters in its down-conversion mixing stages as well as matched filter processing in the detection stage. Although the RF preselector is much wider

than the subsequent IF filtering stages, it serves the valuable function of rejecting strong signal interference that might otherwise cause intermodulation products or other spurious responses in nonlinear components of the receiver. This has benefit for mitigating EMI as well as high-power EA. Another specific application is in image rejection, discussed below.

12.8.2 Image Rejection

Radar receivers generally perform multiple frequency conversions as the received signals are translated from RF to IF and finally to baseband for detection. A jamming signal that is nominally outside the radar waveform bandwidth may end up being translated to the same IF as the target return if it is at an image frequency of one of the mixers. The image frequency for a mixer is located an equal distance from the LO as the desired signal, but in the opposite direction. For example, if the RF signal frequency is $f_{RF} = 8.3$ GHz, and the LO frequency is $f_{LO} = 8.2$ GHz, then the image frequency is $f_{IMAGE} = 8.1$ GHz. The mixer is a nonlinear device that produces many mixing products; among them are the RF input minus the LO reference (the desired IF signal in this example) and the LO minus the RF (the undesired image in this example).

There are two consequences of poor image rejection against EA. The first and most obvious is that the radar is allowing the jamming energy to interfere with the otherwise spectrally separated target signal. The second is that an erroneous angle track error may result if the radar is tracking an SSJ using an image jamming technique. Image jamming exploits the fact that the monopulse error signal derived from the image frequency is in the opposite sense as that derived from the desired frequency. To gain insight into this, recall that the monopulse error signal is the complex ratio of the Δ and Σ channels that includes their relative phase, β , as shown in Equation 12.18. The IF signals produced from the mixing process for the RF and image frequencies are

$$S_{RF} = e^{j(2\pi f_{RF}t + \beta)} e^{-j(2\pi f_{LO}t)} = e^{j2\pi f_{IF}t} e^{j\beta} \quad (12.22)$$

$$S_{IMAGE} = e^{-j(2\pi f_{IMAGE}t + \beta)} e^{j(2\pi f_{LO}t)} = e^{j2\pi f_{IF}t} e^{-j\beta} \quad (12.23)$$

where it is assumed that $f_{RF} = f_{LO} + f_{IF}$ and $f_{IMAGE} = f_{LO} - f_{IF}$, as in the previous example. One can see that the sign of the monopulse phase, β , is different for the image frequency than the desired signal frequency, thereby producing an erroneous error signal.

One method of achieving good image rejection is through RF preselection, described in the preceding section. A complication arises for RF preselection in the case of wideband, frequency agile radars in that the image frequency may be within the tunable bandwidth of the radar. In such a case, a fixed-frequency RF preselector cannot be used. Instead, either a switchable bank of preselectors or a tunable RF filter might be considered; these may prove costly and lossy, however. Another approach is to employ a high first IF. The first IF refers to the difference between the RF and the LO in the first down-conversion stage. If the IF is more than half the tunable RF bandwidth of the radar, the image frequency can be rejected by a fixed, broadband, RF filter that covers the entire tunable RF band of the radar.

Standard receiver design methods typically achieve excellent image rejection. The jammer would need extremely high JSR in order to overcome the radar's image rejection and still compete with the target return. In addition, the jammer must either have precise knowledge of the LO frequency or it must spread its energy over a bandwidth corresponding to the uncertainty in the LO frequency, resulting in power dilution. As a result, jammers are neither likely to attempt nor succeed with image jamming techniques against modern radar.

12.8.3 Wide Instantaneous Dynamic Range

Wide instantaneous dynamic range (IDR) helps prevent receiver saturation while reducing or eliminating the need for AGC. The IDR is the difference between the thermal noise and the strongest signal that can be accurately processed at a given time. Receiver components such as amplifiers, mixers, and limiters can potentially saturate at high input signal levels, although in many current radar systems, it is the ADC that limits IDR. When a component is in saturation, spurious intermodulation products are generated, corrupting the input spectrum. This can occur when strong signals such as clutter, EMI, or EA mix with each other or with target returns. A small-signal suppression effect may also occur whereby the power of a small (desired) signal is further reduced relative to that of a large (undesired) signal at the output of the saturated device.

Automatic gain control is used to partially compensate for limited IDR by adding attenuation in the receiver path to keep the incoming signal within the linear region of operation. The use of AGC creates opportunities for the jammer, however. Two examples are AGC capture and AGC deception. The AGC capture effect is used in conjunction with range and velocity gate stealing techniques. The radar AGC is set based on the amplitude of the deceptive jammer's power before the jammer walks the range or velocity track gate off the target skin return. The radar AGC adds attenuation because of the high JSR of the jammer thereby pushing the skin return down into the noise. As the jammer signal separates from the target in the range or velocity dimension, the target signal is too far down in the noise to be redetected. An instantaneous AGC (IAGC), which can vary the AGC from range bin to range bin, may help mitigate this effect against RGPO [8].

The AGC deception technique introduces angle errors through very fast, high-power, on-off blinking. The AGC attenuation is set based on the average power of the blinking jamming because the AGC time constant is long compared to the jammer blink period. As a result, the AGC attenuation is either too small when the jammer is on or too large when the jammer is off: the radar is alternately saturated and starved, never getting an accurate signal. Against conical scan based angle tracking systems, this AM strips off the scan modulation of the return signal. Against monopulse tracking systems, the jamming distorts the complex difference-to-sum voltage ratio by forcing the signal into nonlinear regions of the receiver where the phase and gain matching characteristics of the monopulse channels are not valid (jammer on) or by forcing low signals levels to be contaminated by thermal noise (jammer off). In all such examples of counter-AGC jamming, the most effective EP solution for the radar is wide IDR.

12.8.4 Notch Filtering

Notch filtering is potentially useful in rejecting narrowband interference when the radar is employing wide or medium bandwidth pulse compression waveforms. Narrowband spot noise or even a CW tone will be decorrelated during pulse compression processing causing the interference to be spread in range. This results in a post-compression interference effect that is comparable to a broadband noise waveform whose bandwidth is approximately matched to that of the radar.

The notch filtering must be performed before any pulse compression processing is done. Initially, this would suggest a series of tunable (or selectable) RF notch filters prior to the first IF conversion. It is often not practical to employ tunable RF notch filters or a large bank of fixed-tuned RF notch filters, however, due to their cost, size, insertion loss,

and low-Q. An alternative approach is to have a dedicated IF stage in which the LO is tuned to place the narrowband interference within a fixed notch filter at the IF. The same LO could be used to translate the remaining signal back to the previous frequency to resume normal signal processing if desired. Several such stages can potentially be cascaded for multiple notch generation if necessary. The narrowband interference power must be below the input third-order intercept point of the mixer, in order to prevent saturation effects.

Ideally, pulse compression is performed entirely through digital signal processing. In this case, optimized notch filters or band-pass filters can be implemented digitally before the pulse compression process occurs, assuming the jamming has not previously saturated the receiver. Notch filtering has negative side effects such as signal loss, elevated range sidelobe levels, and decreased range resolution.

12.8.5 Wideband Limiting

Wideband limiting, also known as a Dicke-Fix, has been used, primarily in older systems, to reduce the effects of intermittent in-band interference in receivers [2,3,10]. A swept noise jammer or a blinking jammer may be in the radar IF band for only a fraction of the time, but the interference may still appear to be continuous if the in-band revisit rate is fast enough. This is because the band-limiting IF filter rings, effectively stretching out the input signal by the impulse response time of the filter, nominally the inverse of the bandwidth. For example, suppose that the final IF filter is 2 MHz and the jammer is in-band for only 100 ns. The output of the filter would persist for a time of approximately 600 ns: 100 ns for the jammer time in band, plus 500 ns for the approximate impulse response time of the filter (the inverse of 2 MHz). If the jammer returns in band before the ringing dies out, it will effectively be continuously present at the detector. The jammer achieves a desired average power level through a combination of high peak power and a low duty factor. The objective of wideband limiting is to limit the peak power of the jammer during its intermittent in-band intervals, thereby also reducing the average power. The limiting action definitely affects the signal quality, but the result may still be preferable to the jamming without limiting.

The limiting cannot be applied after the final IF filter, otherwise the signal will be clipped 100% of the time due to the filter's stretching effect. For a swept noise jammer that is continually within the RF pass band of the radar receiver, limiting also cannot be done immediately after the RF filter because the jammer is always in-band, and the clipping would again occur 100% of the time. Therefore, in the case of swept noise, the limiting is performed at some intermediate stage, following an IF filter whose bandwidth is somewhere between the above extremes.

As an example of how to select such a filter, suppose the jammer has an instantaneous spot noise bandwidth of B_j and sweeps at a rate of ρ (frequency extent divided by sweep time), and the radar IF filter preceding the limiter has a bandwidth of B_{LIM} . The time that the jammer is in-band as it sweeps through the filter is $t_{j,in} = (B_j + B_{LIM})/\rho$. To a first approximation, the jammer output of the filter persists for a time equal to the input time plus the filter impulse response time (nominally $1/B_{LIM}$):

$$t_{j,out} = (B_j + B_{LIM})/\rho + B_{LIM}^{-1} \quad (12.24)$$

A possible choice for the filter bandwidth, B_{LIM} , is one that minimizes the amount of time of the jammer duration after the filter, thereby minimizing the amount of clipping damage

to the desired signal caused by the limiting. This is found by setting the partial derivative of $t_{j,out}$ with respect to B_{LIM} equal to zero (finding the minimum) and then solving for B_{LIM} . Using this approach, the optimum bandwidth is $B_{LIM} = \sqrt{\rho}$. Although this EP approach is more common in older radar designs than in modern designs, it is instructive in that it provides insight into the radar-jammer interaction mechanism in the receiver.

12.8.6 Guard Band Channel

Guard band channels can be used to recognize the presence of noise jamming in radars that do not have Doppler processing [2,3]. Pulsed noise in particular, as might be generated by a noise cover pulse jammer, would have the effect of filling all the range cells corresponding to the duration of the pulse. Without the benefit of Doppler spectral analysis, such a return might be indistinguishable from extended clutter or chaff. The radar can distinguish such noise from extended clutter by using additional receiver channels with frequency bands that straddle the primary radar receiver band. Extended clutter will reside only at the radar frequency and hence will not produce detections in the guard bands. In contrast, a noise jammer will often produce energy in these adjacent bands as well. This allows recognition of the presence of jamming and may lead to subsequent implementation of additional EP.

12.8.7 Noncoherent Detection Channel

A noncoherent detection channel can be used in a pulse compression radar to detect and track noise cover pulse jamming [3]. This type of jamming, in which the noise pulse is generated by a jammer platform during the time it is being illuminated by a radar pulse, may escape detection by other EA-sensing functions in the radar that are designed to detect continuous noise. The NCP is likely to have fairly high JNR to mask target returns following pulse compression. The high JNR allows detection of the jammer in a noncoherent receiver channel with an acceptable combination of probability of detection and probability of false alarm. A range estimate of the jammer platform can be made by tracking the leading edge of the noise pulse, assuming that the rise time is short and the NCP is triggered at the jammer by the arrival of the radar pulse. The recognition of NCP jamming also provides the radar the opportunity to perform angle track-on-jam using the strong jammer signal as a beacon from which to derive its angle estimates, as described in Section 12.10.6.

12.8.8 Least Jammed Frequency Selection

A radar receiver that is capable of characterizing the EA noise spectrum can be used in conjunction with frequency agility to select a transmit frequency that has the least amount of jamming interference [2,6,10]. Spot noise and multiple spot noise techniques are obvious examples where this would apply, but the approach can potentially be used against BN jamming as well. The BN jamming spectrum received by a radar is not necessarily uniform, and therefore the radar may achieve significant EP benefit by locating gaps, nulls, or local minima in the noise spectrum. The BN jammer spectrum may not be uniform for several reasons. First, the jammer may not be able to cover the entire radar bandwidth with a single waveform generator or even a single amplifier but rather must cover the frequency

extent in contiguous subbands. The regions of overlap of these subbands may have lower power due to band roll-off, or there may be gaps between the subbands. Second, the jammer transmit amplifier, transmission line, and antenna may have frequency-dependent gain, phase, and voltage standing wave ratio (VSWR) properties that create an uneven power spectrum. Third, multipath may cause constructive or destructive interference that is frequency-dependent. Finally, the radar antenna sidelobe pattern may have frequency-dependent variations in the direction of the jammer.

The ability to sense the EA interference spectrum before the next transmission is sometimes referred to as a sniff mode or sniffer [2,3,6]. Radar schedulers may introduce an inherent lag between the time that the environment is sensed and the time the selected frequency is actually used; thus, there is a risk that the assessment made during the sniff mode is stale when the least jammed frequency is selected. The sniff function is usually scheduled to occur at a time in the radar timeline during which there is no expected return from a target or clutter in the environment. A question that may arise in implementing this technique for ESA radars is whether to adjust the antenna beam-steering command with frequency as the EA spectrum is characterized. This may give a more accurate representation of the likely noise level at the frequency ultimately selected, but the beam-steering action may prove impractical.

12.9 | SIGNAL PROCESSOR-BASED EP

The signal processor accepts the receiver ADC samples and performs high-speed operations, including pulse compression, temporal filtering, Doppler integration, power centroiding, track error signal derivation, and threshold detection. The signal processor EP techniques are applied within the time scale of a CPI. Applicable EP techniques include detection and mitigation of both noise and deceptive EA, in some cases enabling subsequent EP to be performed in the data processor.

12.9.1 Total Energy Detector

The total energy detector (TED) is a conceptually simple method for detecting the presence of noise or coherent masking. The signal processor computes the energy in a large number of range or Doppler cells and compares the total to some threshold representative of an EA-free environment. If the total energy exceeds the threshold, then jamming is suspected. Ideally, the energy samples are taken from a range-Doppler region that excludes high clutter returns; otherwise the total energy computation may be biased by extended clutter returns and could lead to a false declaration of jamming. The TED does not inherently distinguish between unintended EMI and EA. However, subsequent logic in the data processor might be able to correlate high-TED occurrences with specific beam positions or multiple observations over time to help determine whether the interference is EMI or EA. The TED detection may be used to generate jam strobes, described in Section 12.10.6.

12.9.2 Alternate Constant False Alarm Rate Detection

Radars with CFAR detection employ detection thresholds that vary based on estimates of the interference energy (such as noise and clutter) contained in a particular detection cell, referred to as a cell under test (CUT). The CFAR detection process is described in [8].

The estimate of the noise energy in the detection cell of interest is based on an average of the measured energy in a sample group of range or Doppler cells, called reference cells, which are usually in the vicinity of the CUT but which necessarily exclude the CUT and surrounding guard cells to avoid being biased by the signal itself. The detection threshold is set at some level above the noise estimate (e.g., 15 dB) to keep the false alarm rate at some tolerable level. The CFAR mechanism prevents a flood of false alarms from occurring under conditions of external interference such as noise jamming by raising the detection threshold. Although the elevated detection threshold desensitizes the radar, this is preferable to a high false alarm rate, which might effectively prevent detection of true targets in any event by overloading the radar. A CFAR algorithm may be optimized for interference with specific statistical properties; white Gaussian noise is a logical choice as it corresponds to receiver thermal noise. Interference due to jamming may have different statistical properties than those for which the radar CFAR algorithm is optimized, however; this may cause the threshold to be improperly set, leading to either additional false alarms or additional loss of sensitivity.

Generally it is not desirable for the radar to be desensitized by noise jamming without also detecting that jamming is present: such a design would essentially make targets with noise jamming invisible to radar. The radar sniff mode and the TED, described already, are well suited for detecting strong noise jammers but may not be able to detect subtler EA techniques such as range bin masking or narrowband Doppler noise, which limit the EA interference to a very narrow range or Doppler region in the vicinity of the target.

One way of potentially improving CFAR performance against localized EA masking is to use a double sliding window CFAR approach, as depicted in Figure 12-12. Instead of a single pair of noise estimation windows surrounding the cell under test, the radar uses two pairs. The estimates derived from the inner and outer pairs are compared, and the lower of the two noise estimates is used to determine the detection threshold. A bin masking technique that focuses the EA energy in one of the regions may miss the other. This will make it more likely that the threshold is not biased higher by the EA. The result will be detection of an extended region that includes both target and EA, making their presence at least visible to the radar.

The CFAR detection performance may also suffer from the presence of strong targets or clutter discretely in the noise sample region. This might artificially elevate the CFAR threshold and potentially mask targets. The same effect potentially results from EA false targets. A CFAR algorithm that manages to exclude strong outliers from the noise estimate might perform better under such circumstances than conventional cell-averaging CFAR. Two examples of this are ordered statistic CFAR (OS-CFAR) and censored cell CFAR (CC-CFAR) [6]. Both approaches first rank-order the amplitude measurements in the reference cells, arranging them in a sequence of lowest amplitude to highest amplitude.

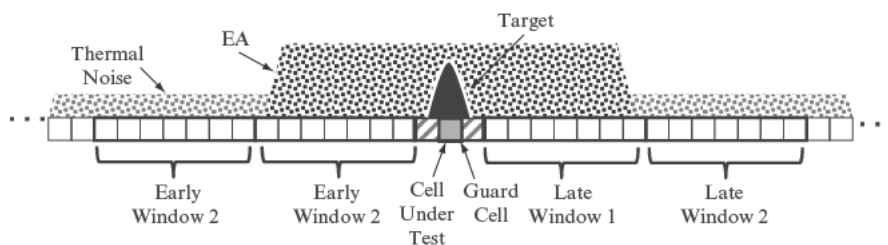


FIGURE 12-12 ■
Double-sliding
window CFAR with
localized EA
masking.

The OS-CFAR computes the mean interference based on a subset of the sampled cells that exclude the k cells with the highest values, with the value of k being a design parameter. The CC-CFAR picks a specific cell of the rank-ordered sequence as representative of the true, EA-free noise, analogous to picking the median instead of the mean. These algorithms help ensure that the true target is detected, although the false targets (or clutter discretized) may also be detected in the process.

12.9.3 Angle Variance

A high variance in the angle measurements made under conditions of main beam noise jamming may be indicative of multiple jammers in the main beam. The relative amplitude and phase of the two noise jammers will vary in a random fashion because their noise waveform generators are independent. The angle measurement on average will be near the power centroid of the multiple jammers, but individual measurements will vary considerably due to random variations of their relative amplitude and phase. For the case in which two noise jammers are blinking, either synchronously or asynchronously, such that there are intervals during which only one jammer is on at a time, the distribution will tend to have a more bimodal characteristic to it. In contrast, a single jammer with a high JNR will be characterized by a narrow distribution of angle measurements around a single angle. Noise jammers will cover many if not all of the range-Doppler cells used for detection. The signal processor can form the monopulse ratio on a large number of the range-Doppler cells in a single CPI and compute the statistical properties of the noise jamming angle distribution. The angle variance is more reliable for high-JNR conditions because the angle measurements are less influenced by thermal noise.

12.9.4 Data Editing

Radars whose CPIs are long compared to the ADC sampling interval may be able to employ data editing, or excising, to minimize effects of intermittent jamming. This concept is similar in spirit to wideband limiting, except in this case, the jammer sweep time is long compared to the impulse response time of the band-limiting IF filter. As a result, a jammer that is either pulsed or sweeps in and out of the radar band will manifest itself as an intermittent on-off modulation during the radar CPI. Ordinarily, the radar signal processor coherently integrates the jammer spikes over the CPI, and the resulting effect is comparable to a continuous jammer with an average power equal to the intermittent jammer's peak power multiplied by its duty factor. The data editing technique is applicable if the jammer has a low duty factor and high peak power relative to the target signal, thermal noise, and clutter.

The signal processor must recognize unnaturally large increases and decreases in power over the CPI and then excise these data intervals from the data before performing the coherent signal integration. There is an accompanying loss in signal energy and degradation of range or Doppler image quality. The signal energy in the detection cell decreases by a factor $(1 - d_j)$ for a given jamming duty cycle d_j within the CPI; this should be more than offset, however, by a much greater decrease in jamming energy under intended conditions. One of the likely consequences of the data editing is elevated range or Doppler sidelobes. These sidelobes may increase the amount the clutter interference in the target detection cell in clutter-intensive environments. This cost must be weighed against the potential benefit of reducing the EA impact. Signal processing algorithms that partially restore

short intervals of missing or corrupted data may be considered to reduce the damage done to the underlying data.

Another example of how a data editing type of technique might be applied is in the case of a narrowband noise jammer against a wideband pulse compression waveform, analogous to notch-filtering in the receiver. A radar that is able to sample the entire waveform bandwidth with its ADC and perform the pulse compression digitally in the signal processor may be able to form a digital notch filter to remove the corrupted part of the spectrum; alternatively, it may be able to form contiguous digital band-pass filters and pass only the interference-free portion of the spectrum. The fraction of signal lost is approximately equal to the jammer noise bandwidth divided by the radar waveform bandwidth. This approach will not work, however, for a radar that employs stretch processing to achieve wideband LFM pulse compression. In that case, the deramping function, typically at the first IF stage, will spread the narrowband noise in frequency, making subsequent notch filtering impossible. (A fallback approach for this condition might be to temporally excise the portion of the deramped signal during which the EA interference falls within the IF bandwidth.)

12.9.5 Guard Gates

Guard gates are used to recognize the presence of either RGPO or VGPO track deception techniques [3,4]. During a RGPO (or VGPO) event, the radar range (or Doppler) track gate will tend to follow the deceptive jammer signal and move away from the target skin return because the jammer signal is much stronger and therefore dominates the range (or Doppler) track error signal. The guard gates consist of two range (or Doppler) cells that are on either side of the track gate, as illustrated in Figure 12-13. As the track gate moves away from the target, the target eventually appears in one of the guard gates. This raises a flag indicating the possibility of a RGPO (or VGPO). If the radar concludes that this is an RGPO (or VGPO) in process, possibly through the confirmation of some other EP logic, then it can initiate a snap-back and bring the track gate back to the skin, whose position is presumed to be that of the activated guard gate. This technique requires that the radar receiver have sufficient dynamic range to avoid AGC capture, as described in Section 12.8.3; otherwise, the weaker target return would be attenuated and may not have sufficient SNR to be detected in the guard gate.

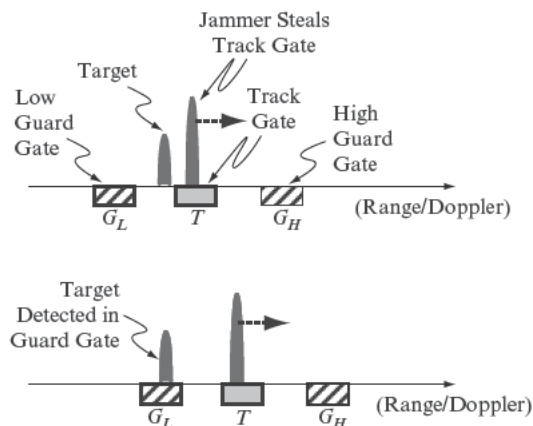
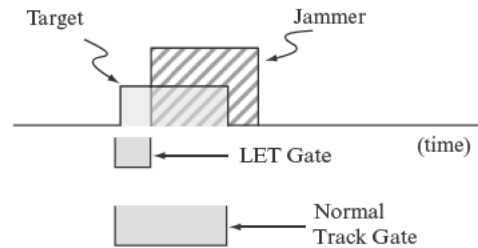


FIGURE 12-13 ■
Guard gate used to detect gate stealing.

FIGURE 12-14 ■
Leading-edge track
concept.



Ideally, the signal processor has the ability to detect and measure multiple targets in the beam at the same time, rather than rely on a few guard gates in range or Doppler. Maintaining track on multiple potential targets in the beam would prevent the radar from being deprived of a potential signal to track in the event that some of the false targets are dropped or are used in combination with angle deception techniques. Electronically scanned array radars are especially well suited for this as they are largely intended to perform multiple target tracking.

12.9.6 Leading-Edge Track

Leading-edge track (LET) helps counter coherent repeaters that have significant throughput delay [2,6]. The radar uses a range track gate that is much narrower than the radar pulse width and is biased toward the leading edge of the target return, as shown in Figure 12-14. Coherent repeaters with sufficiently long throughput delay will fall outside of the LET range gate and will not be able to affect the range, Doppler, or angle tracking functions of the radar. This provides a benefit similar to the narrow pulse and pulse compression waveforms in that the target return is isolated from the slightly delayed jammer. In the case of the LET, however, the radar does not use a matched filter to accomplish the detection. The radar IF bandwidth must be wide enough and the ADC sample rate high enough to accommodate the narrow LET track gate if LET is to be performed in the signal processor. For example, if the LET gate is 200 ns, then the IF bandwidth must be at least 5 MHz (1/200ns). Otherwise, the LET would have to include an analog gating function in the receiver. Leading edge track may also be used against NCP jammers. The delay in this case is due to detection, measurement, and response time of the jammer.

The primary drawback with LET is the loss in SNR, corresponding to the reduction in available pulse width. Leading-edge tracking is therefore limited to conditions in which the target SNR is high, such as for close-range or high-RCS target engagements. Coherent repeaters can potentially defeat LET by generating a coherent cover pulse over the leading edge; this is certainly possible for DRFM-based jammers with PRI tracking operating against repetitive, range-ambiguous waveforms such as those used for pulse-Doppler.

12.10 | DATA PROCESSOR-BASED EP

The data processor accepts detection reports and position and feature measurements from the signal processor and performs target tracking and higher-level logic operations. Electronic protection techniques at this stage allow detection of false targets as well as detection and tracking of gate stealing and noise techniques.

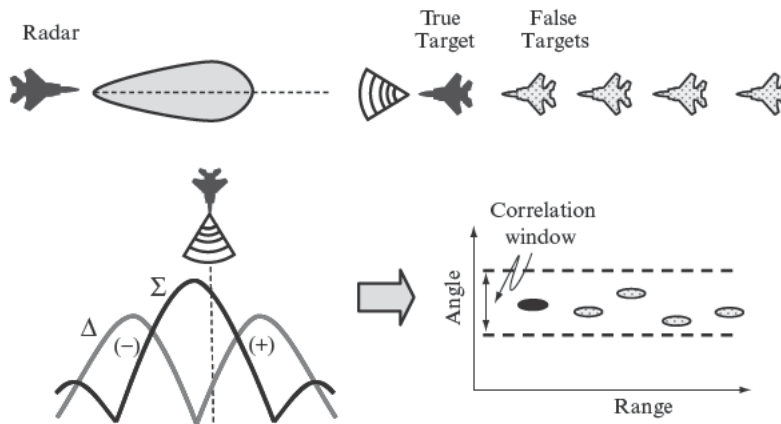


FIGURE 12-15 ■
Angle alignment of
false targets.

12.10.1 Angle Alignment Correlation

Angle alignment correlation can be used to detect the presence of range or Doppler false targets in the main beam. Figure 12-15 illustrates the concept. Multiple detections in the main beam whose azimuth and elevation measurements all fall within a narrow angular window provide a tip-off of a potential false target generator. The correlation window is based on the measurement and track estimation accuracy of the radar. The longer the alignment condition persists as the radar and target platform geometry changes, the more likely it is that a false target jammer is present. In the case of multiple RFTs, the radar can select the nearest one as the true target provided that the jammer does not have the ability to generate a false target up-range from its own position. This assumption may not always be valid, however, as coherent repeaters with PRI-tracking and PRI-delay capability can potentially place targets up-range against radar waveforms such as pulse-Doppler.

12.10.2 Radar Cross Section Statistics

The RCS statistics over time of a target being tracked can be used to determine whether the target is true or false. Complex targets such as aircraft have multiple scatterers that are unresolved in range and Doppler and that undergo varying amounts of constructive and destructive interference as the engagement geometry changes. The Swerling target types described in [8] constitute the most common method of classifying targets based on their RCS fluctuation characteristics. A jammer false target that does not include AM to replicate RCS fluctuations may be distinguishable from a true target. The radar data processor may collect and analyze multiple RCS measurements over time, say, several seconds, compensating for expected R^{-4} received power variation.

One disadvantage from the jammer's perspective in implementing an RCS-like AM is that it may require operation of the linear region of the transmit amplifier rather than in compression. This reduces the average power delivered by the transmitter, which may be undesirable from a cost or performance standpoint. The jammer transmitter might not even be designed to run in a linear mode and must run in compression at all times. Performing the AM after the HPA in this case is also an unattractive option due to the requirement for a high-power variable attenuator.

12.10.3 Radar Cross Section Step Detection

A large, sudden rise in the apparent RCS (amplitude) of a target being tracked can be used as an alert of a potential track gate stealing technique [9]. Jammers attempting to execute RGPO or VGPO techniques must have sufficient JSR to steal the track gate away from the skin. When the jammer turns on, the apparent RCS of the target will suddenly increase by approximately the JSR and remain at the higher average level. The algorithm used to recognize the step must be immune to false alarms triggered by normal target fluctuations. Thus the data processor would compare multiple-CPI averages taken before the apparent step with those taken after the step.

12.10.4 Range-Doppler Track Comparison

A pulse-Doppler radar can potentially detect the presence of RGPO, VGPO, or false targets by comparing the estimates of range rate, or radial velocity, obtained independently through the range and Doppler track filters [6]. This is sometimes referred to as an $\dot{R} - v$ (“R-dot-v”) comparison. The range tracker estimate of range rate, denoted v_R , is derived from the time rate of change of the estimated range. The Doppler, or velocity, tracker estimate of range rate, denoted v_V , is obtained directly from the Doppler frequency estimate, f_D :

$$v_V = -f_D \frac{c}{2f_c} \quad (12.25)$$

where f_c is the radar carrier frequency. The negative sign accounts for a positive Doppler corresponding to a negative range rate; that is, the range is decreasing with time. The absolute value of the difference between the two velocity estimates, Δv_{RV} , will be less than some maximum error tolerance Δv_{MAX} for a true target return. (This assumes that velocity ambiguities have already been resolved in the Doppler track filter.) The error tolerance Δv_{MAX} accounts for the individual error margins for the v_R and v_V estimates. A deceptive jammer will be detected if its range and Doppler programs are not accurately coordinated, causing Δv_{RV} to exceed Δv_{MAX} .

A jammer may be capable of executing a deceptive program in only one of the two dimensions—range or Doppler—due to cost or technology limitations. Before the widespread use of DRFMs, coherent RF memories were somewhat limited in capability (such as bandwidth and memory depth) and were more costly to implement than Doppler modulators. Doppler radars that did not perform the $\dot{R} - v$ comparison could be defeated with VGPO alone, and therefore jammers could afford to have limited or no RGPO capability.

Even for a coordinated range-velocity technique, the $\dot{R} - v$ comparison might still be able to detect a deceptive jammer. The jammer generates the desired velocity offset relative to the jamming platform, v_{FT} , by applying a corresponding Doppler shift, $f_{D,FT}$:

$$f_{D,FT} = -\frac{2v_{FT}f_j}{c} \quad (12.26)$$

where f_j is the value of the radar RF used by the jammer in its velocity-to-Doppler conversion. Although a repeater does not require explicit knowledge of the radar RF to coherently repeat it, it must have some estimate of the RF to generate a Doppler shift consistent with the intended velocity offset, as Doppler is proportional to RF ($f_D = 2vf_c/c$).

The radar interprets the velocity of the jammer's Doppler offset as

$$v_{V,FT} = -\frac{c}{2f_c} f_{D,FT} = v_{FT} \frac{f_j}{f_c} \quad (12.27)$$

In other words, the apparent range rate offset of the false target as observed by the Doppler tracker is different by a factor of f_j/f_c from the jammer's intention. An error in the absolute velocity of a false target or VGPO is not necessarily a problem for the jammer if the radar does not compare the range and Doppler filter estimates. However, if the radar does perform the $\dot{R} - v$ comparison, the error may be detected.

In contrast, the jammer's range-rate offset will appear as intended in the radar's range tracker because it is based on a time variation of the jammer coherent delay, and the conversion of range offset to coherent delay does not require any measurement or assumption of radar parameters: as per Equation 12.1, $t_{\text{delay}} = 2R_{\text{offset}}/c$. The difference between the range and Doppler tracker velocity estimates now includes the discrepancy between the jammer's coherent delay and Doppler programs:

$$\Delta v_{FT} = v_{FT} - v_{FT} \frac{f_j}{f_c} = v_{FT} \frac{f_c - f_j}{f_c} = v_{FT} \frac{\Delta f_j}{f_c} \quad (12.28)$$

where Δf_j is the difference between actual and assumed RF. Thus, a 10% error in the jammer's assumption of the radar RF will produce a disagreement between the two track filters of 10% of the intended range-rate offset relative to the jammer platform.

As an example, suppose that the jammer intends to impart an apparent 300-m/s difference in radial velocity between its own platform and the range-velocity false target, and the jammer assumes the radar RF to be 9 GHz. The Doppler offset it applies is then 18,000 Hz. Suppose, however, that the radar RF is actually 10 GHz, and the jammer is simply assuming a midband frequency within an 8–10 GHz channel. In this case there is a 10% error in the assumed RF, and there would be an inconsistency in radial velocity of approximately 10%, or 30 m/s, between the range and Doppler track filters. This corresponds to a difference in Doppler of 2000 Hz. One would expect this to far exceed any measurement errors in the radar, as typical Doppler filter widths of pulse-Doppler radars are less than half this amount.

A radar with frequency agility can attempt to force the above condition against jammers with poor real-time frequency measurement capability [6]. Many jammers are capable of measuring frequency to an accuracy of better than a few MHz on a single pulse. The processing logic needed to incorporate real-time frequency updates in the Doppler offset used for VGPO is very straightforward (a simple multiply) and therefore the VGPO program can be updated with each change in radar RF. A radar with the ability to transmit simultaneously on two or more frequencies, as described in Section 12.7.8, makes the task of the jammer much more difficult. To avoid being detected by range-Doppler comparison against such a radar, the jammer must apply an independent Doppler modulation to each radar frequency, accomplished through channelization in the receiver or technique generator.

12.10.5 Track Filter Acceleration Limit

The radar range, Doppler, azimuth, and elevation track filters can be designed with acceleration limits to reject EA false targets that exhibit kinematically unrealistic behavior [6,9]. It is in the interest of the radar to keep the track filter gains from being unnecessarily

high in order to avoid overreaction to measurement noise or target maneuvers. In contrast, it is in the interest of the jammer to complete any range, velocity, or angle track deception techniques in as short a time period as possible. If the jammer's overly energetic pull-off results in a false target whose acceleration is considered impossible for the class of target being engaged, the radar can reject the deceptive technique and restore track on the target skin return. For example, if the jammer's goal is to achieve a 1-km RGS in 5 seconds or less, it would require at least an 8-g pull-off acceleration (where g is the acceleration due to gravity, or 9.8 m/s^2). If the radar's filter is only designed for 4-g targets, the pull-off would fail. The jammer would instead have to achieve the 1-km pull-off in approximately 7 seconds, potentially increasing its vulnerability.

One way the acceleration limit might take effect is that the radar track filter is simply too sluggish to keep up with the jammer false target and never gets drawn away from the true target skin return. Another possibility is that the radar recognizes the high acceleration, declares a potential EA condition, and attempts to establish separate tracks on the true and false targets, employing a higher gain filter for the latter. The radar can still potentially use the false target to obtain a high-SNR estimate of range, velocity, or angle if any of these parameters is not included in the EA track deception technique. Deliberately reducing the gain of a track filter to counteract deceptive EA is a more aggressive EP approach but is also highly risky because the radar is susceptible to losing track on a real target that is conducting a maneuver.

12.10.6 Track-on-Jam

Track-on-jam, angle-on-jam, and home-on-jam (HOJ) are all variations on a common theme: the radar uses the jammer's transmission to perform angle tracking when the jammer has otherwise denied accurate range and Doppler estimates of the target. The specific case of an HOJ mode refers to a missile seeker using the jammer signal for guidance. Though these techniques are most associated with countering noise jammers, the basic concept applies to deceptive jammers as well. An example is the radar using the angle estimate from an SSJ false target or gate stealer, but ignoring the false range and Doppler information. In the case of a continuous noise jammer, the radar is not obligated to radiate toward the jammer, but can simply operate in a passive listening mode.

Jammer angle information is sometimes used to generate a jam strobe on the radar display, indicating a line of bearing of a detected jammer. Although target range may be undetermined, the presence and direction of the jamming provides useful situation awareness. Radars that operate in a networked configuration with other, spatially displaced radars can potentially use the jammer strobes to triangulate on the jammer, thereby reducing the range uncertainty.

12.10.7 Angle Gating

Angle gating is a method whereby the radar rejects very large angle track error signals and extrapolates its angle track estimate until acceptable error levels resume [3]. The large angle error signal may result from the abrupt application of an angle deception technique or it may be due to target glint. Under such conditions, the radar suddenly receives an angle track error that is indicative of a large leap in angular position made by the target in track. Since such behavior does not correspond to a natural target, except for possible brief instances of glint, the radar gates-out the large error signal and temporarily coasts

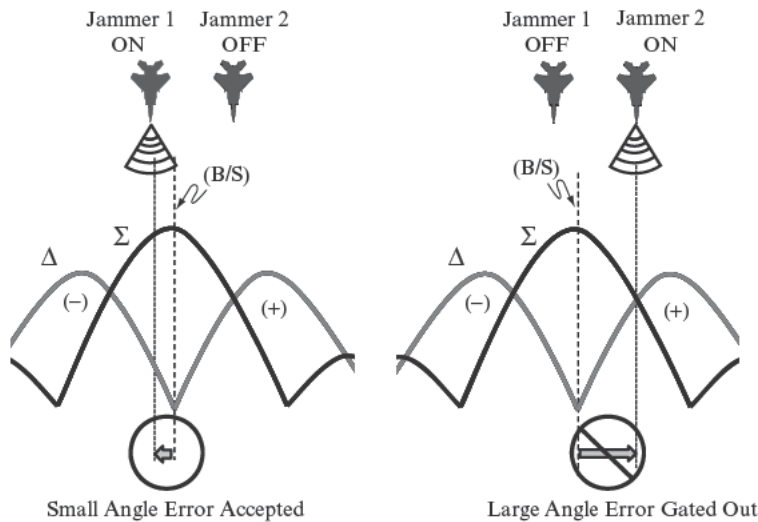


FIGURE 12-16 ■
Angle gating of large
angle errors from
blinking noise
jammer.

in angle. If the source of the anomalous angle error is of sufficiently short duration, the radar angle track can be resumed without having suffered the disruption of a large error impulse to its tracking estimate.

One possible application of this technique is in the case of two blinking noise jammers in the main beam, as illustrated in Figure 12-16. If the blinking results in conditions where only one jammer is on at a given time, the radar track filter will have difficulty settling on either jammer, as it is alternately driven first to one and then to the other. In the angle gating approach, the radar might attempt to establish a stable angle track on one of the jammers and ignore the influence of the other by gating out the high errors. Alternatively, multiple noise-jammer track files might be established in the radar where the radar can associate the different angle error measurements to the separate tracks and maintain a stable angle estimate on each platform.

12.11 | SUMMARY

As is evident from this chapter, radar EW includes a wide range of methods by which adversary EA and ES systems can degrade or exploit radar operation. The EA effects primarily consist of masking and deception and provide screening of targets through either support jamming or self-protection jamming deployments. Jamming is generated through either coherent or noncoherent EA system architectures: the former can preserve the phase information of the radar, the latter cannot. There are numerous EA techniques, operating in the range, Doppler, or angle dimensions of the radar, including noise, false targets, and track gate stealing. Table 12-3 lists the EA techniques mentioned briefly in this chapter, along with their intended effect, jammer type, and jammer role. There are numerous variations to forming such a list, but this provides a representative example of the diversity of EA options. The techniques are grouped in terms of general objective rather than alphabetically.

An important measure of jammer effectiveness is the JSR, which provides a measure of how strong the jamming is relative to the target in the radar detection cell. Expressions

TABLE 12-3 ■ Summary of EA Techniques

EA Technique		Primary Effect		Jammer Type		Jammer Role	
Name	Abbrev.	Masking	Deception	Noncoherent	Coherent	Support (SOJ, SIJ, ESJ)	Self-Protect (SSJ)
Spot noise	SN	x		x		x	x
Barrage noise	BN	x		x		x	x
Blinking spot noise	BSN	x		x			x
Blinking barrage noise	BBN	x		x		x	x
Multiple spot noise	MSN	x		x			
Responsive spot noise	RSN	x		x		x	x
Noise cover pulse	NCP	x		x	x	x	x
CW range denial	CWRD	x			x	x	x
Doppler noise	DN	x			x	x	x
Range bin masking	RBM	x			x		x
Velocity bin masking	VBM	x			x		x
AGC capture	AGCC	x			x		x
Range false targets	RFT		x		x	x	x
Velocity false targets	VFT		x		x	x	x
Sidelobe false targets	SLFT		x		x	x	
Range gate pull-off	RGPO		x		x		x
Velocity gate pull-off	VGPO		x		x		x
Inverse gain	IG		x	x	x		x
Swept AM	SAM		x	x	x		x
AGC deception	AGCD		x	x	x		x
Image jamming	IMJ		x	x	x		x
Chaff burst	CHF		x				x
Towed RF decoy	TRD		x	x	x		x
Expendable decoy	EXPD		x	x	x		x
Terrain bounce	TB		x	x	x		x
Jammed chaff	JAFF		x	x	x		x
Cross-pol	XPOL		x		x		x
Cross-eye	XEYE		x		x		x
Coop. blinking noise	CBN		x	x	x	x	x

for the JSR of a noise jammer, saturated coherent repeater, and linear coherent repeater are, respectively,

$$JSR = \frac{P_j G_{j,tx} G_{rj} d_j 4\pi R^4}{R_j^2 B_j P_t G_t G_r \sigma n \tau} \quad (12.29)$$

$$JSR_{SAT} = \frac{P_j G_{j,tx} G_{rj} d_j 4\pi R^4}{R_j^2 P_t G_t G_r \sigma n_j} \quad (12.30)$$

$$JSR_{LIN} = \frac{G_{tj} G_{rj} G_{j,rx} G_{j,tx} g_{j,el} \lambda^2 R^4 d_j}{G_t G_r 4\pi \sigma R_j^4 n_j} \quad (12.31)$$

Electronic protection capabilities can be implemented within any functional subsystem in the radar: the antenna, transmitter, exciter, receiver, signal processor, and data processor. The EP techniques allow filtering, cancelling, resolving, attenuating, or distinguishing a jamming signal relative to the desired target. The EP can be applied in the spatial, temporal, spectral, or polarization domains. Table 12-4 provides a summary of the

TABLE 12-4 ■ Summary of EP Techniques

Subsystem	EP Capability	Examples of Applicable EA	Specific EA Assumption
Antenna	Low sidelobes	BN, SN, RSN, SLFT	Sidelobe
Antenna	Sector blanking	RSN, SLFT	Needs main lobe to respond
Antenna	Sidelobe blanking	SLFT	Low duty factor
Antenna	Sidelobe cancellation	BN, SN, RSN	Sidelobe noise
Antenna	Main lobe cancellation	BN, SN, RSN	Main lobe ESJ noise
Antenna	Narrow beamwidth	BN, SN, RSN, CBN	Main lobe ESJ noise
Antenna	Random beam scan	SLFT	Scan-synchronous
Antenna	Scan on receive only	IG	AM-based angle pull-off
Antenna	Lobe on receive only	IG	AM-based angle pull-off
Antenna	Monopulse angle	IG, SAM	AM-based angle pull-off
Antenna	Scan with compensation	IG, SAM	AM-based angle pull-off
Antenna	Low cross pol antenna	XPOL	Polarization based EA
Transmitter	High peak power	BN, SN, RSN, TB, XEYE, XPOL	High-JSR dependent
Transmitter	Transmit power reduction	RSN, NCP, ARM, RGPO, SLFT	EW receiver-dependent
Transmitter	Emission control	RSN, NCP, ARM	EW receiver-dependent
Exciter	Wide-pulse waveform	BN, SN, RSN	Moderate JSR
Exciter	Multiple-pulse waveform	BN, SN, RSN	Moderate JSR
Exciter	Burnthrough waveform	BN, SN, RSN	Moderate JSR
Exciter	Narrow pulse	RSN, NCP, RGPO, VGPO	Delayed-response SSJ
Exciter	Pulse compression	RSN, NCP, RGPO, VGPO	Delayed-response SSJ
Exciter	Doppler resolution waveform	CHF, EXPD	Different velocity than target
Exciter	Frequency diversity	BN, SN	Non-frequency sensing
Exciter	Frequency agility	BN, SN, RSN, NCP, VGPO, RGPO	Slow-frequency sensing
Exciter	Multiple simultaneous frequency	BN, SN, RSN, SLFT, RGPO, VGPO	Limited channels or sidelobe
Exciter	PRF variation	RSN, RGPO, NCP, RFT, SLFT	PRI-based identification or track
Receiver	RF preselection	BN, MSN, IMJ	Off-frequency, high-JSR
Receiver	Image rejection	BN, MSN, IMJ	Image frequency, high-JSR
Receiver	Wide IDR	AGCD, AGCC	High-JSR, rapid on/off
Receiver	Instantaneous AGC	RGPO, AGCC	High-JSR RGPO
Receiver	Notch filtering	SN, MSN	Narrowband
Receiver	Wideband limiting	BN, BBN, BSN	Fast swept/blinking
Receiver	Guard band channel	BN, SN, MSN, RSN	Wideband noise
Receiver	Noncoherent detection channel	NCP, BBN, BSN	High-JNR pulsed
Receiver	Least jammed frequency selection	BN, SN, MSN	Uneven spectrum
Signal Proc.	Total energy detector	BN, SN, RSN, DN	High-JNR noise
Signal Proc.	Double window CFAR	DN, RBM, VBM, RFT, VFT	Localized noise
Signal Proc.	Ordered statistic CFAR, censored cell CFAR	RFT, VFT	Strong discretes
Signal Proc.	Data editing	BN, BBN, BSN	Slow swept/blinking
Signal Proc.	Angle variance	BN, SN, NCP	Multiple main lobe noise
Signal Proc.	Guard gates	RGPO, VGPO, EXPD	Gate-stealing
Signal Proc.	Leading edge track	RSN, NCP, RGPO, VGPO	Delayed response SSJ
Data Proc.	Angle alignment correlation	RFT, VFT, RGPO, VGPO	Main lobe SSJ
Data Proc.	RCS statistics	RFT, VFT, RGPO, VGPO	Constant power
Data Proc.	RCS step detection	RGPO, VGPO	High-JSR
Data Proc.	Range-Doppler track comparison	RFT, VFT, RGPO, VGPO	Uncoordinated range/Doppler
Data Proc.	Track filter acceleration limit	RGPO, VGPO, XPOL, CBN	High-acceleration
Data Proc.	Angle-on-jam, track-on-jam, home-on-jam	BN, SN, RSN, DN, RGPO, VGPO	High-JNR SSJ
Data Proc.	Angle gating	CBN, XEYE, XPOL	Large, intermittent angle error

EP techniques discussed in this chapter as well as examples of applicable EA techniques and associated specific assumptions for the EA. (The EA technique abbreviations used in Table 12-4 are found in Table 12-3.)

Although it may appear that we have covered a large number of EP options, in a sense we have only scratched the surface. The discussion in this chapter is by no means intended to be comprehensive but rather is meant to provide an introductory glimpse into a broad and multifaceted topic. Not only are there many additional EP techniques besides the ones described, but of those that are described we have touched on only a few of the key issues determining applicability and performance. Hopefully, however, the reader has come away with an understanding of some of the fundamental concepts of radar EP as well as an appreciation for the diversity and intensity of the ongoing EW battle for control of the electromagnetic spectrum.

12.12 | FURTHER READING

An encyclopedic type overview of EW is provided in the classic, two-volume set by Leroy Van Brunt [2,3]. Although decades old, these books still provide excellent insight into the breadth of EA and EP (then called ECM and ECCM) from the perspective of a true practitioner.

A clear description of key EP concepts as applicable to airborne pulse-Doppler radar is found in [6]. A broad overview of radar EW is also provided in [1,7].

For those interested in an historical perspective, two books trace the use of EW from World War II through the 1960s and then from the 1960s through the 1990s:

- Price, A., *Instruments of Darkness (The History of Electronic Warfare)*, Peninsula Publishing, Los Altos, CA, 1987.
- Price, A., *War in the Fourth Dimension (US Electronic Warfare, from the Vietnam War to the Present)*, Stackpole Books, Mechanicsburg, PA, 2001.

12.13 | REFERENCES

- [1] Schleher, D.C., *Electronic Warfare in the Information Age*, Artech House, Norwood, MA, 1999.
- [2] Van Brunt, L.B., *Applied ECM*, vol. 1, EW Engineering, Dunn Loring, VA, 1978.
- [3] Van Brunt, L.B., *Applied ECM*, vol. 2, EW Engineering, Dunn Loring, VA, 1982.
- [4] Van Brunt, L.B., *The Glossary of Electronic Warfare*, EW Engineering, Dunn Loring, VA, 1984.
- [5] Eaves, J. and Reedy, E., Eds., *Principles of Modern Radar*, Chapman and Hall, New York, 1987.
- [6] Morris, G. and Harkness, L., Eds., *Airborne Pulsed Doppler Radar*, 2d ed., Artech House, Norwood, MA, 1996.
- [7] Neri, F., *Introduction to Electronic Defense Systems*, 2d ed., SciTech Publishing, Raleigh, NC, 2006.
- [8] Richards, M.A., Scheer, J.A., and Holm, W.A., Eds., *Principles of Modern Radar, Vol. 1: Basic Principles*, SciTech Publishing, Raleigh, NC, 2010.

- [9] Stimson, G.W., *Introduction to Airborne Radar*, 2d ed., Part VIII, Radar in Electronic Warfare, SciTech Publishing, Raleigh, NC, 1998.
- [10] Skolnik, M., Ed., *Radar Handbook*, 2d ed., Chapter 9, Electronic Counter-Countermeasures, McGraw-Hill, New York, 1990.
- [11] Adamy, D., *EW 101: A First Course in Electronic Warfare*, Artech House, Norwood, MA, 2001.

12.14 | PROBLEMS

1. A coherent repeater is programmed to produce a RFT that is 1200 m farther in range from the radar than the jammer platform itself. What is the time delay that must be generated by the jammer's coherent RF memory to achieve this apparent range delay? What is the rate of variation of this delay, in μs per second, if the RFT is to be perceived as moving farther away in range at a rate of 300 m/s?
2. What is the Doppler offset required to make the RFT motion in Problem 1 appear kinematically consistent to a radar at 10 GHz RF?
3. You are assigned to select one of two jamming pods for a specific mission of countering a known, fixed frequency radar from a SOJ platform. The radar has a waveform bandwidth of 5 MHz. Your two choices of EA pods are a 10-kW, 500-MHz barrage noise jammer, and a 1-kW, 20-MHz responsive spot noise jammer. Which jammer provides the greater power spectral density and by how much (in dB)?
4. In Problem 3, assume the radar operates at 10 GHz and has an antenna gain toward the jammer of -5 dBi and a system noise figure of 6 dB. What is the JNR of the selected jammer if it is operating in the sidelobe at a range of 50 km with an ERP at the radar polarization of 40 dBW?
5. Equation 12.10 shows the JSR for a noise jammer from a SSJ platform against a monostatic radar. What would the equation be for a bistatic radar, where R_t is the range to the target from the transmitting radar platform, and R_r is the range to the target from receiving radar platform?
6. A coherent SSJ operating as a linear repeater must generate a 20-dB JSR RGPO technique over a 20-m^2 target against a 10-GHz radar. How much net repeater throughput gain must the jammer have to accomplish this assuming all of its power is going in a single range-Doppler cell? Assuming the jammer HPA has a peak power of 20 W and is transmitting 5 W at range 100 km, at what range would the HPA be likely to go into compression, changing the jammer from a linear repeater to a saturated repeater?
7. A 10-GHz radar has a peak transmitter power of 10 kW, a main lobe antenna gain of 40 dBi, and a sidelobe gain of approximately -10 dBi. What system sensitivity, in dBmi, would the jammer need to detect the radar in the main lobe from 100 km? What system sensitivity would it need to detect the radar in the sidelobe from 50 km?
8. An SSJ repeater has a delayed response of 167 ns relative to the leading edge of the target skin return. What is the minimum range resolution needed by the radar, either through a narrow pulse or pulse compression to be able to resolve the target return from the jammer? What approximate waveform bandwidth does this correspond to?

PART IV

Post-Processing Considerations

CHAPTER 13 Introduction to Radar Polarimetry

CHAPTER 14 Automatic Target Recognition

CHAPTER 15 Multitarget, Multisensor Tracking

Introduction to Radar Polarimetry

Krishna Naishadham

Chapter Outline

13.1	Introduction.....	589
13.2	Polarization.....	594
13.3	Scattering Matrix.....	601
13.4	Radar Applications of Polarimetry.....	611
13.5	Measurement of the Scattering Matrix.....	618
13.6	Summary.....	622
13.7	Further Reading.....	622
13.8	References.....	623
13.9	Problems.....	626

13.1 | INTRODUCTION

Electromagnetic (EM) waves are characterized by oscillatory motion of the field vector in a plane perpendicular to the direction of wave propagation. Unlike acoustic waves, which are unpolarized due to vibrations only in the direction of propagation, the instantaneous vector orientation of the electric field over a plane transverse to the propagation vector makes an EM wave polarized. The time-dependent behavior of the electric field vector, in general describing an ellipse in the transverse plane, plays an important role in the interaction of EM waves with material bodies (commonly referred to as targets in radar nomenclature). Radar polarimetry is concerned with control of the coherent polarization properties of radar signals and the extraction of target (or scatterer) properties from the behavior of its back-scattered returns within the microwave spectral bands. Radar polarimetry exploits the vector nature of EM fields and has become an indispensable tool in radar systems. Wave polarization plays a very important role in target identification [1–6]. The polarization behavior of the reflectance and scattering from, for example, foliage, ocean, earth, rain, and chaff has been extensively used in remote sensing and clutter mitigation [7–11]. The propagation medium itself can transform the polarization state of an incident wave due to variation in the index of refraction (or permittivity, permeability, and conductivity), and the EM wave is said to be repolarized. As an example, the ionosphere has various effects on spaceborne imaging radar, one of which is Faraday rotation of the polarization plane. With a fully polarimetric radar, the amount of Faraday rotation in a specified magnetic field can be measured over point or distributed targets whose signature has a defined and known

orientation about the radar line of sight. Polarimetric measurements subject to Faraday rotation can be calibrated using a calibration target, such as a sphere or trihedral corner reflector, which has no cross-polarized return. The effect of Faraday rotation can then be estimated and removed from the measured scattering matrix of an arbitrary scatterer. The reader is referred to Section 10.4 in [11] and the references therein for a detailed discussion on the estimation and correction of Faraday rotation.

When the wave strikes an object such as a radar target, characteristic information about its material composition, shape, and orientation can be obtained from measurement of the polarimetric scattering matrix in a suitably defined orthogonal coordinate basis [12–22]. The scattering matrix accounts for all information contained in the polarized target return under the conditions of the given radar encounter, such as frequency, waveform, target observation angle, and presence of clutter. Measurement of the complete matrix requires using dual polarized antennas for transmission and reception to characterize co-polarized and cross-polarized received states for each of the two orthogonal states transmitted. Due to cost and complexity, most practical radar systems measure only the partial matrix, thereby trading some information on the target for a less expensive system that operates more quickly while providing the information that is most often used. Polarimetry deals with the exploitation of coherent polarization properties of EM waves and their interaction with the scattering object, observed through complex received voltage or power or radar cross section (RCS) measurements, from which the scattering matrix can be derived. Various radar applications of polarimetry involving target identification, feature recognition through radar imaging, and remote sensing of terrain and weather patterns have been extensively developed over the years [1–11, 23–26].

The foundations of radar polarimetry were laid down in the early 1950s by Kennaugh [12,13]. Kennaugh's reports [13] document the pioneering research on radar polarimetry conducted by his research group between September 16, 1949, and October 1, 1954. Kennaugh first introduced the concept of scattering matrix in polarimetry and demonstrated that there exist characteristic polarization states for which the radar receives minimum/maximum power. This work laid the foundation for radar polarimetry and formed a rigorous theoretical basis behind the design of early (pre-1960) civilian and defense polarimetric radar systems worldwide. Kennaugh introduced the theory to estimate the polarization states for optimal reception of a coherent EM wave. The optimization procedure for the channel power leads to an eigenvalue equation for the radar scattering matrix, whose diagonalization using unitary matrix transformation yields the characteristic polarization states for which the radar receives minimum or maximum power in the reciprocal, monostatic case. Of particular importance are the eigen-polarization states that produce nulls in the cross-polarization and co-polarization channels. Huynen [14] generalized Kennaugh's work to time-varying targets and showed that a time-varying power-averaged target return (e.g., chaff cloud, ocean surface) is characterized by nine significant polarization parameters, whereas a fixed target can be described by five parameters. This resulted in the pioneering polarization fork concept, which enabled easy visualization of the null polarization states on the Poincaré sphere. He introduced the first target decomposition theorem and showed that the time-averaged target can be decomposed into an effective single fixed (time-invariant) target that represents the ensemble, and a residue target or N -target, which represents target noise or depolarizing scattering behavior at the higher frequencies. Huynen applied the phenomenological theory of radar targets to scattering from rough surfaces, such as terrain and sea, and rigorously accounted for the depolarizing behavior of such objects. This work provided the basis for accurate characterization of land, sea,

and precipitation clutter using improved analysis and decomposition of the polarization states [11]. Boerner modified and extended Kennaugh's method of determining optimal polarizations to non-reciprocal and bistatic (or partially polarized) cases [27–30].

It is evident that polarization is an important parameter to consider in radar target identification comprising the three phases of target detection, discrimination, and recognition. This chapter presents an introduction to radar polarimetry emphasizing the polarization behavior of monochromatic waves as described by the scattering matrix. After a brief review of polarization theory, Stokes parameters and scattering matrices are introduced and explained using examples on canonical targets. The distinction between Kennaugh's and Huynen's formulations of the target scattering matrix (partially polarized case) is discussed. To demonstrate the application of radar polarimetry in radar imaging, the scattering matrix measurements of a cone with grooves (resembling a missile reentry vehicle) are processed to extract key features such as specular scattering and edge diffraction contributions from range-Doppler images, and the usefulness of polarimetric signatures to discriminate between coated and uncoated bodies is discussed. The polarimetric behavior of precipitation clutter, sea clutter, and ground clutter is discussed briefly.

13.1.1 Organization

Section 13.2 reviews EM wave polarization theory. In Section 13.2.1, linear, circular, and elliptical polarization is discussed, culminating in the derivation of geometrical parameters of the polarization ellipse. Section 13.2.2 presents the Stokes parameters and their relationship to the Poincaré sphere. Section 13.3 reviews the polarimetric scattering matrix formulation in linear and circular polarization bases and examines simplification of the matrix for rotationally symmetric targets. The Sinclair or Jones matrix is introduced in Section 13.3.1. In Section 13.3.2, Huynen's formulation of the optimal polarization states for signal reception are reviewed, culminating in a discussion of the polarization fork concept. Partial polarization described by the Stokes reflection or the Mueller matrix, and its relationship to the Kennaugh matrix, are the subject of Section 13.3.3. Polarimetric applications of scattering matrix theory to target detection and feature extraction, which are useful in the discrimination phase, are discussed in Section 13.4, focusing on static-range measured data for a cone with a spherical nose tip. New results are presented that demonstrate the stark contrast of polarization-dependent returns between uncoated (metallic) and dielectric-coated objects. The types of polarization most suitable for reducing the effect of ground, sea, and precipitation clutter are briefly reviewed. Measurement of the polarization scattering matrix is discussed in Section 13.5, considering a high-level description of the necessary equipment and instrumentation setup. The chapter concludes with a summary, followed by suggestions for further reading on advanced polarimetric techniques.

13.1.2 Key Points

- Fundamentals of wave polarization (linear, circular, elliptical)
- Geometrical parameters of the polarization ellipse, such as tilt angle, ellipticity angle, and the rotational sense
- Polarization bases for linear, circular, and elliptical states and transformation between them
- Stokes parameters for fully and partially polarized waves and description of the Poincaré sphere

- Scattering matrix definition for radar targets, Sinclair and Jones formulations, transformation of the matrix by change of bases between orthogonal polarization states
- Eigendecomposition of the scattering matrix using unitary transformations
- Determination of optimal transmit/receive polarization states that maximize or minimize backscattered power
- Null polarization states and Huynen polarization fork
- Stokes reflection or Mueller matrix for partial polarization and its relationship with Kennaugh matrix
- Instrumentation setup for scattering matrix measurement (block diagram description)
- Impact of polarimetry on radar target detection
- Extraction of radar target features from polarimetric signature data
- Polarimetric behavior of rain, sea, and ground clutter

13.1.3 List of Symbols

The following lists the variable names found within this chapter:

r, θ, ϕ = radius, elevation and azimuth angles in the spherical coordinate system

f = frequency (Hz)

$\omega = 2\pi f$ = radian frequency (rad/s)

k = wavenumber (1/m)

λ = wavelength (m)

c = velocity of light (m/s)

t = time (s)

E = electric field intensity (V/m)

δ = phase difference between orthogonal field components (rad)

P = polarization ratio

p = polarization efficiency

\mathbf{e}_t = transmitter antenna polarization

\mathbf{e}_r = receive antenna polarization

$\hat{\mathbf{e}}_t, \hat{\mathbf{e}}_r$: transmit and receive polarization state (unit) vectors

CO-POL: Co-polarized transmit and receive states

CROSS-POL: Cross-polarized transmit and receive states

Huynen Fork Parameters:

m = maximum received signal for the optimal transmit polarization state

ψ = orientation angle or tilt angle (radar coordinate system)

γ = target polarizability angle

ν = target skip angle (characterizes multiple bounces)

τ = ellipticity angle of the polarization ellipse

AR = axial ratio of the polarization ellipse

(g_0, g_1, g_2, g_3) = Stokes parameters

$g(\mathbf{a}) = (g_0, g_1, g_2, g_3) = (g_0, \mathbf{g})$ = Stokes vector of transmit antenna with polarization \mathbf{a}

$h(\mathbf{b}) = (h_0, h_1, h_2, h_3) = (h_0, \mathbf{h})$ = Stokes vector of receive antenna with polarization \mathbf{b}

d = degree of polarization

M: Mueller matrix

K: Kennaugh matrix

S : target scattering matrix
 S_{SMR} : target scattering matrix with relative phase
 S_{SMA} : target scattering matrix with absolute phase
 S_d : diagonalized target scattering matrix
 U : unitary matrix
 M_0 : Mueller matrix (fixed symmetrical target)
 N : Noise or residue (non-symmetrical) target
 J, K, L : Pauli spin (rotation) matrices
 V_{rt} = votage at terminals of the receive antenna
 σ = radar cross section

13.1.4 Acronyms

Commonly used acronyms used in this chapter include the following:

AR	axial ratio
BSA	back-scatter alignment
COHO	coherent oscillator
CP	circular polarization
CW	continuous wave
DBF	digital beamformer
EM	electromagnetic
EP	elliptical polarization
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques
FSA	forward-scatter alignment
H	Hermitian
HH	co-polarized (horizontal) received signal for horizontal Tx polarization
HV	cross-polarized (horizontal) received signal for vertical Tx polarization
IF	intermediate frequency
LCP	left circular polarization
LO	local oscillator
LP	linear polarization
LSM	linear restricted scattering matrix
PNA	programmable network analyzer
PO	physical optics
RCP	right circular polarization;
RCS	radar cross section
RF	radio frequency
Rx	receiver
SAR	Synthetic aperture radar
SMA	scattering matrix with absolute phase
SMR	scattering matrix with relative phase
SNR	signal-to-noise ratio
SSM	state space method
STALO	stable local oscillator
T	transpose
Tx	transmitter
VV	co-polarized (vertical) received signal for vertical Tx polarization
VH	cross-polarized (vertical) received signal for horizontal Tx polarization

13.2 POLARIZATION

13.2.1 Fundamentals of Wave Polarization

The fundamentals of EM wave polarization are introduced in this section. Linear, circular, and elliptical polarizations are considered, and the parameters of the polarization ellipse, namely, axial ratio, tilt angle, and sense of polarization, are discussed. The display of polarization states on the Poincaré sphere is explained, and the description of polarization in terms of Stokes parameters, commonly employed in radar polarimetry, is presented.

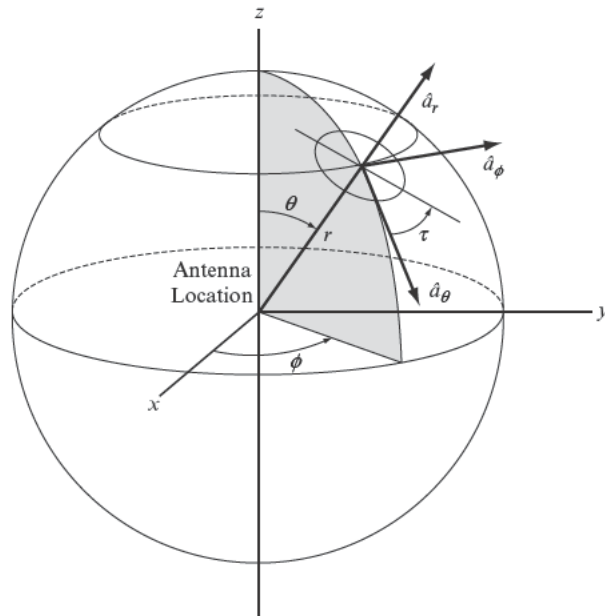
Consider a wave radiated by an antenna in a spherical coordinate system. In general, the electric field will have orthogonal tangential components, E_θ, E_ϕ , oriented perpendicular to the (radial) direction of wave propagation, as shown in Figure 13-1 [31]. Polarization refers to the locus of the instantaneous electric field vector tip at a fixed point on the spherical equiphase surface and changes with the elevation angle, θ , and azimuth angle, ϕ . If the tip traces out a straight line as a function of time, the wave is linearly polarized, circularly polarized if the locus is a circle, and elliptically polarized if the tip describes an ellipse. In general, the polarization of the wave at a point in space is elliptical in nature, characterized by two parameters, the axial ratio and the tilt angle.

Without loss of generality, we consider a plane wave propagating along the $\hat{k} = \hat{z}$ direction with its electric field decomposed into two orthogonal components perpendicular to the direction of propagation. Thus, for a given location in space, we write

$$E(z, t) = \hat{x}E_x + \hat{y}E_y = \hat{x}E_{0x} \cos(\omega t - kz) + \hat{y}E_{0y} \cos(\omega t - kz + \delta) \quad (13.1)$$

where E_{0x} and E_{0y} denote amplitudes, ω is the angular frequency, $k = \omega/c$ is the wave number in free space, c denotes the wave velocity, and δ corresponds to the phase difference between the y -component and the x -component (δ can be positive or negative). Polarization remains constant in any plane perpendicular to the direction of propagation.

FIGURE 13-1 ■
Orientation of the polarization ellipse and the field vectors on a sphere. (From Balanis [31]. With permission.)



Therefore, we examine the locus described by the tip of the electric field vector \mathbf{E} in the plane $z = 0$, as time changes, for different values of E_{0x} , E_{0y} and δ .

13.2.1.1 Linear Polarization

When the phase difference $\delta = 2n\pi$, n an integer, the two components are in phase, and we have $E_x/E_{0x} = E_y/E_{0y}$. Thus, the wave is linearly polarized with its locus traversing a positive slope in the first and third quadrants (Figure 13-2a). When the phase difference is an odd multiple of 2π , that is, $\delta = (2n + 1)\pi$, we have $E_x/E_{0x} = -E_y/E_{0y}$, and the straight line locus traverses negative slope in second and fourth quadrants (Figure 13-2b).

13.2.1.2 Circular Polarization

When the amplitudes of the two components in (13.1) are equal ($E_{0x} = E_{0y} = E_0$) and the phase difference $\delta = \pm\pi/2$, the electric field is given by

$$\mathbf{E}(t) = \hat{\mathbf{x}} E_0 \cos(\omega t) \mp \hat{\mathbf{y}} E_0 \sin(\omega t) \quad (13.2)$$

It can be seen that as ωt increases from 0 to 2π , the tip of the electric field vector traverses a circle, and thus the wave is circularly polarized. When $\delta = +\pi/2$, the tip rotates counterclockwise with the observer looking into the direction of propagation (out of the page in Figure 13-2), and the wave is said to be left-hand polarized. Equivalently, if the left-hand thumb points to the direction of propagation while the fingers of the left hand curl in the direction of rotation of the electric vector, the wave is left-hand polarized. Conversely, the wave is right-hand polarized when the sense is described by right-hand thumb and fingers, as can be verified for the case $\delta = -\pi/2$. For example, the loci for these two senses of circular polarization are depicted in Figure 13-2c (RCP) and Figure 13-2d (LCP). It may be seen that the tip rotates along the circle of radius E_0 with a constant angular rate ω .

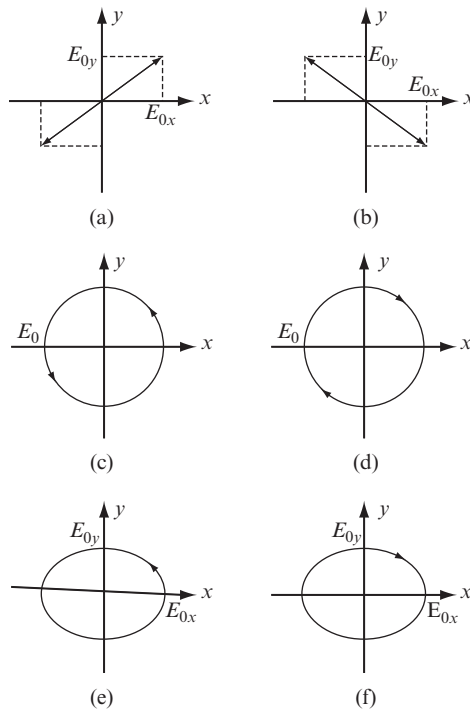


FIGURE 13-2 ■ Polarization states in a plane perpendicular to direction of propagation. (a) linear polarization (LP) with positive slope. (b) LP with negative slope. (c) Right circular polarization (RCP). (d) Left circular polarization (LCP). (e) Right-handed elliptical polarization (EP). (f) Left-handed EP. The propagation direction is out of the page.

13.2.1.3 Elliptical Polarization

In general, the tip of the electric field vector in (13.1) traces out an ellipse for arbitrary values of E_{0x}, E_{0y} and δ , and the wave is elliptically polarized [20]. Consider the case when the amplitudes of the two components in (13.1) are not equal (say, $E_{0x} > E_{0y}$) and the phase difference $\delta = \pm\pi/2$. The electric field is now given by

$$E(t) = \hat{x} E_{0x} \cos(\omega t) \mp \hat{y} E_{0y} \sin(\omega t) \quad (13.3)$$

It can be readily seen that the wave is elliptically polarized, with the major and minor axes of the ellipse aligned with x- and y-axes, respectively (Figures 13-2e and 13-2f). As we trace the ellipse with increasing ωt over one time period, we infer left-handed EP for $\delta = +\pi/2$, and right-handed EP for $\delta = -\pi/2$. If the phase difference is not an odd multiple of $\pi/2$, the polarization ellipse will be tilted from the Cartesian coordinate axes, as shown in Figure 13-3. Next, we derive the axial ratio, defined as the ratio of major to minor axis, and the orientation or tilt angle, ψ , relative to the x-axis. The axial ratio is also specified by the ellipticity angle, τ (see Figure 13-3), within the range $-\pi/4 \leq \tau \leq \pi/4$. These two angles, ψ and τ , completely specify the polarization ellipse, with the sense of polarization derived from the algebraic sign on the angle τ . For future reference, it is also convenient to define ellipticity in terms of the auxiliary angle $\gamma = \tan^{-1}(E_{0y}/E_{0x})$, $0 \leq \gamma \leq \pi/2$. The polarization ellipse can be alternatively specified by either the angle pair (γ, δ) derived from the field components (see (13.1)), or the angle pair (ψ, τ) derived from the ellipse geometry. It will be shown that these two pairs are related.

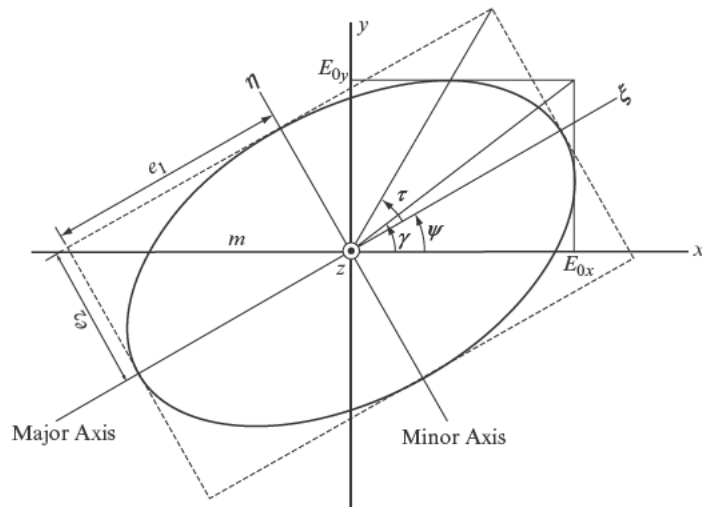
By eliminating ωt from (13.1) we obtain the equation to the ellipse given by

$$\left(\frac{E_x}{E_{0x}}\right)^2 + \left(\frac{E_y}{E_{0y}}\right)^2 - 2\frac{E_x E_y}{E_{0x} E_{0y}} \cos \delta = \sin^2 \delta \quad (13.4)$$

where δ is the phase difference between the x- and y-components. Let us consider the polarization ellipse in a coordinate system rotated from the Cartesian system by the tilt angle ψ , as shown in Figure 13-3. The Cartesian field components in the (ξ, η) coordinate system are given by

$$\begin{aligned} E_x &= E_\xi \cos \psi - E_\eta \sin \psi \\ E_y &= E_\xi \sin \psi + E_\eta \cos \psi \end{aligned} \quad (13.5)$$

FIGURE 13-3 ■
Polarization ellipse.
The propagation
direction is out of
the page.



After substituting (13.5) in (13.4) and simplification, we obtain

$$\frac{E_\xi^2 \cos^2 \psi - E_\xi E_\eta \sin 2\psi + E_\eta^2 \sin^2 \psi}{E_{0x}^2} + \frac{E_\xi^2 \sin^2 \psi + E_\xi E_\eta \sin 2\psi + E_\eta^2 \cos^2 \psi}{E_{0y}^2} - \frac{(E_\xi^2 - E_\eta^2) \sin 2\psi + 2E_\xi E_\eta \cos 2\psi}{E_{0x} E_{0y}} \cos \delta = \sin^2 \delta \quad (13.6)$$

Equation (13.6) is the general solution for the polarization ellipse in the (ξ, η) coordinate system. Since the ξ - and η -axes coincide with the major and minor axes of the ellipse, the sum of the cross-product terms involving $E_\xi E_\eta$ must vanish. Therefore

$$\frac{E_\xi E_\eta \sin 2\psi}{E_{0y}^2} - \frac{E_\xi E_\eta \sin 2\psi}{E_{0x}^2} - \frac{2E_\xi E_\eta \cos 2\psi}{E_{0x} E_{0y}} \cos \delta = 0 \quad (13.7)$$

which may then be solved for the tilt angle ψ using the relation

$$\tan 2\psi = \frac{2E_{0x} E_{0y} \cos \delta}{E_{0x}^2 - E_{0y}^2}, \quad -\frac{\pi}{2} \leq \psi \leq \frac{\pi}{2} \quad (13.8)$$

From Figure 13-3, we may define an auxiliary angle

$$\gamma = \tan^{-1} \frac{E_{0y}}{E_{0x}}, \quad 0 \leq \gamma \leq \frac{\pi}{2} \quad (13.9)$$

and rewrite the tilt angle in (13.8) as

$$\tan 2\psi = \tan 2\gamma \cos \delta, \quad \gamma \neq \frac{\pi}{4} \quad (13.10)$$

If $E_{0x} = E_{0y}$ (or $\gamma = \pi/4$), the tilt angle cannot be calculated from (13.8) or (13.10). Instead, it may be found from the circular polarization ratio (Chapter 3 in [3]).

Next, we derive the axial ratio of the ellipse in Figure 13-3, defined as $AR = e_1/e_2$, in terms of the ellipticity angle

$$\tau = \cot^{-1}(\mp AR), \quad -\frac{\pi}{4} \leq \tau \leq \frac{\pi}{4} \quad (13.11)$$

The tilt angle τ is negative for right-handed and positive for left-handed polarization, respectively. Since AR is positive, it follows that the negative sign in (13.11) leads to right-handed sense, and the positive sign, to left-handed sense. It is left as an exercise to show that the major axis e_1 and the minor axis e_2 satisfy the relations

$$e_1^2 = E_{0x}^2 \cos^2 \psi + E_{0y}^2 \sin^2 \psi + E_{0x} E_{0y} \sin 2\psi \cos \delta \quad (13.12)$$

$$e_2^2 = E_{0x}^2 \sin^2 \psi + E_{0y}^2 \cos^2 \psi - E_{0x} E_{0y} \sin 2\psi \cos \delta \quad (13.13)$$

$$e_1 e_2 = E_{0x} E_{0y} \sin \delta \quad (13.14)$$

Clearly,

$$e_1^2 + e_2^2 = E_{0x}^2 + E_{0y}^2 \quad (13.15)$$

Using the trigonometric identity $\sin 2\tau = (2 \tan \tau)/(1 + \tan^2 \tau)$, it follows from (13.11), (13.14), and (13.15) that

$$\sin 2\tau = \frac{2e_1 e_2}{e_1^2 + e_2^2} = \frac{2E_{0x} E_{0y} \sin \delta}{E_{0x}^2 + E_{0y}^2} \quad (13.16)$$

Alternatively, the ellipticity angle can be written succinctly in terms of the auxiliary angle γ (see (13.9)) as

$$\sin 2\tau = \sin 2\gamma \sin \delta \quad (13.17)$$

In summary, the polarization ellipse is completely specified by (a) the tilt angle given by (13.8), (b) the ellipticity angle in (13.16), and (c) the sense of rotation specified by the sign on the ellipticity angle as determined from (13.11).

The elliptically polarized field components in the rotated (ξ, η) coordinate system (see Figure 13-3) may be written in complex vector form as

$$\mathbf{E}(\xi, \eta) = \begin{bmatrix} E_\xi \\ E_\eta \end{bmatrix} = m \begin{bmatrix} \cos \tau \\ j \sin \tau \end{bmatrix} e^{j\alpha_0} \quad (13.18)$$

where $m = \sqrt{e_1^2 + e_2^2}$ is the field amplitude, the angle τ is negative for right-handed EP and positive for left-handed EP (see (13.11)), and α_0 is an arbitrary reference phase angle common to both E_ξ and E_η . Converting the phasor fields in (13.18) into real time-dependent form using $\mathbf{E}(t) = \text{Re}[(E_\xi \hat{\xi} + E_\eta \hat{\eta})e^{j\omega t}]$, it can be readily verified that the time-dependent fields satisfy the equation to the ellipse with semi-major axis, e_1 , and semi-minor axis, e_2 , given by

$$\frac{E_\xi^2(t)}{e_1^2} + \frac{E_\eta^2(t)}{e_2^2} = 1 \quad (13.19)$$

The complex field components in the (x, y) coordinate system follow from (13.5) and (13.18) as

$$\mathbf{E}(x, y) = \begin{bmatrix} E_x \\ E_y \end{bmatrix} = m e^{j\alpha_0} \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \begin{bmatrix} \cos \tau \\ j \sin \tau \end{bmatrix} = m e^{j\alpha_0} \hat{\mathbf{h}}(x, y) \quad (13.20)$$

where the complex unit vector $\hat{\mathbf{h}}$ represents the polarization state, and is known as the Jones vector [3]. Equation (13.20) defines the field components in the (x, y) coordinate system entirely in terms of the tilt angle and the ellipticity angle and forms the basis for scattering matrix representation of polarization.

13.2.2 Stokes Parameters and Poincaré Sphere

It is convenient to represent polarization of a monochromatic wave in terms of real-valued Stokes parameters that may be derived from the polarization ellipse specified by the tilt angle and the ellipticity angle. Stokes parameters permit an interpretation of various polarizations states as points on the Poincaré sphere. The polarization states can be defined in terms of a Stokes vector, comprising the Stokes parameters, as [14, 32],

$$\mathbf{g} = \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} |E_x|^2 + |E_y|^2 \\ |E_x|^2 - |E_y|^2 \\ 2 \text{Re}(E_x E_y^*) \\ 2 \text{Im}(E_x E_y^*) \end{bmatrix} = g_0 \begin{bmatrix} 1 \\ \cos 2\tau \cos 2\psi \\ \cos 2\tau \sin 2\psi \\ \sin 2\tau \end{bmatrix} \quad (13.21)$$

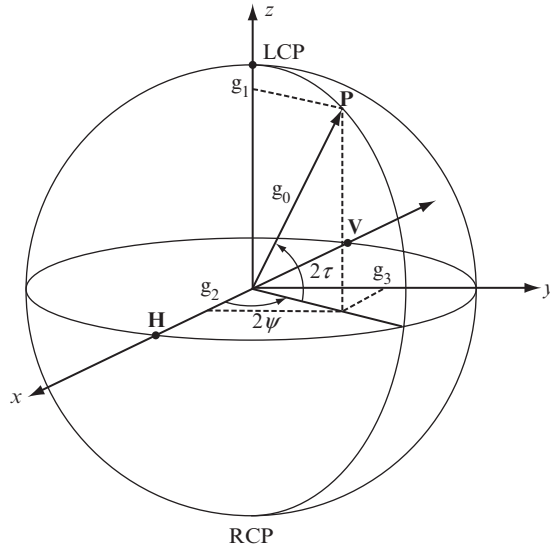


FIGURE 13-4 ■
Poincaré sphere.

where $*$ denotes the complex conjugate, $g_0 = \sqrt{g_1^2 + g_2^2 + g_3^2}$ denotes the total power for a completely polarized wave, and therefore, the parameters g_1, g_2, g_3 are independent. In general, the wave is partially polarized, and the total average power satisfies the condition $g_0 > \sqrt{g_1^2 + g_2^2 + g_3^2}$, and $g_1 = g_2 = g_3 = 0$ for an unpolarized wave [34]. Stokes parameters for partial polarization will be reviewed in Section 3.3. The derivation of the Stokes vector in terms of tilt and ellipticity angles of the polarization ellipse (the last column in (13.21)) is left as an exercise. Equation (13.21) suggests a simple geometrical interpretation of all polarization states by recognizing that g_1, g_2, g_3 are the rectangular components of a point on a sphere with radius g_0 (see Figure 13-4), known as the Poincaré sphere. The elevation and azimuth angles in the spherical coordinate system are given by $\theta = 2\tau$ and $\phi = 2\psi$, respectively. Alternatively, 2ψ and 2τ are the longitude and latitude, respectively, of the point on the sphere. Since positive τ occurs for left-hand polarization and negative τ for right-hand polarization (see (13.11)), the former polarization states map on to the upper hemisphere and the latter to the lower. These locations on the sphere will be reversed if $\exp(-i\omega t)$ time convention is employed. In fact, all corresponding orthogonal polarizations are antipodal on the Poincaré sphere. Some examples will be discussed next.

Case 13.1: Linear Polarization

If $\tau = 0$, the wave is linearly polarized along a ray inclined at the tilt angle ψ . It follows from (13.21) that the Stokes vector is given by

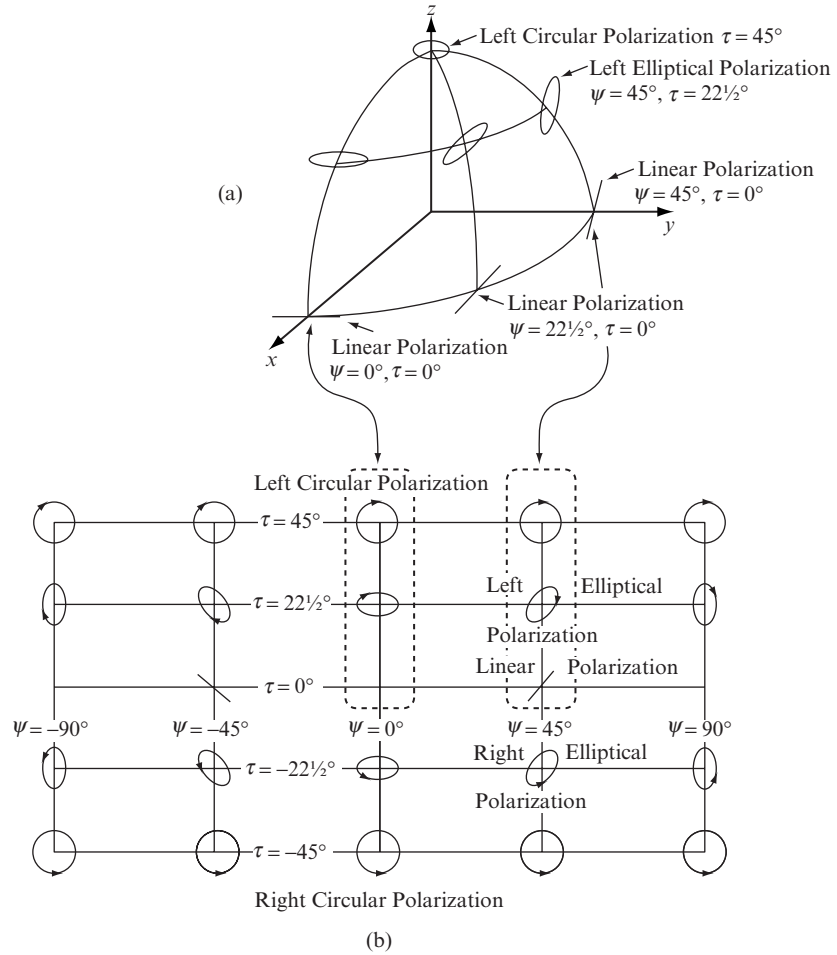
$$\mathbf{g} = g_0 \begin{bmatrix} 1 \\ \cos 2\psi \\ \sin 2\psi \\ 0 \end{bmatrix} \quad (13.22)$$

Thus, any point on the equator of the Poincaré sphere represents a state of linear polarization. For $(\tau, \psi) = (0, 0)$, the wave is horizontally polarized along the x-direction, and the polarization state is located at $(g_1, g_2, g_3) = (g_0, 0, 0)$. Likewise, $(\tau, \psi) = (0, \pi/4)$ results in a polarization state on the y-axis with an orientation at 45° relative to the horizontal,

FIGURE 13-5 ■

(a) Polarization states on one octant of the Poincaré sphere.

(b) Rectangular projection of the Poincaré sphere showing full range of polarization states (From Kraus [32]. With permission.). The propagation direction is out of the page.



and $(\tau, \psi) = (0, \pi/2)$ yields a vertically polarized state located on the negative x-axis. For ease of visualization, a rectangular projection of the various polarization states on the Poincaré sphere is shown in Figure 13-5 [32].

Case 13.2: Circular Polarization

For $\tau = \pm\pi/4$, the wave is circularly polarized and the polarization state is located at the poles of the Poincaré sphere, with the North Pole representing left circular polarization and the South Pole representing right circular polarization. Note that the tilt angle, ψ , is undefined in the case of circular polarization.

Case 13.3: Elliptical Polarization

In general, the wave is elliptically polarized, with the polarization state located at a point on the upper hemisphere for the left-handed sense and at a point on the lower hemisphere for the right-handed sense. The two limiting cases pertain to circular polarization at the poles and linear polarization on the equator. Several elliptical states of polarization are shown in Figure 13-5 by ellipses with appropriate tilt angle ψ and ellipticity angle τ drawn at points on the Poincaré sphere. As an example, for $(\tau, \psi) = (\pi/8, \pi/8)$, we obtain the

left-hand elliptical polarization state at the spherical angles $(\pi/4, \pi/4)$. This ellipse is inclined 22.5° relative to the horizontal, as depicted in Figure 13-5 [32].

As an application of the Poincaré sphere representation, it may be shown that voltage at the terminals of a receiving antenna due to an incident wave of arbitrary polarization (radiated by a transmitting antenna) is given by [27]

$$V = C \cos \frac{MM_a}{2} = C |\hat{\mathbf{h}} \cdot \hat{\mathbf{h}}_a^*| \quad (13.23)$$

where M and M_a are points on the Poincaré sphere corresponding to polarization states $\hat{\mathbf{h}}$ and $\hat{\mathbf{h}}_a$, for the incident wave and the antenna, respectively, MM_a is the angle subtended by great-circle line from the polarization state M to M_a , and C is a constant that depends on the field strength of the wave and the effective aperture of the antenna. In particular, if the angle $MM_a = 0$, the polarization states of the transmit and receive antennas are matched, and the response in (13.23) is maximized. This happens, for example, at the poles of the Poincaré sphere, when two circularly polarized states of the same sense overlap, or on the equator, between two co-polarized linear states. If the angle $MM_a = 180^\circ$, the polarization states are orthogonal, and no signal is received. Thus, for maximum signal at the terminals of the receiving antenna, the transmit and receive antennas must be co-polarized. Conversely, no signal is received if the two antennas are cross-polarized.

It is convenient to calculate the polarization efficiency of a communication link from (13.23) as

$$p = \cos^2 \frac{\alpha}{2}, \alpha = \angle MM_a \quad (13.24)$$

Assuming completely polarized states, the angle α may be calculated in terms of Stokes parameters of the transmit and receive antennas as

$$\cos \alpha = \frac{g_1^t g_1^r + g_2^t g_2^r + g_3^t g_3^r}{g_0^t g_0^r} \quad (13.25)$$

with the superscripts t and r denoting the transmitter and receiver, respectively. Using the polarization efficiency, receive antenna aperture, and the incident power density, we can calculate the net power received [3].

13.3 | SCATTERING MATRIX

In microwave circuit theory, the term ‘scattering matrix’ is used to describe voltage measurements in a multi-port network. However, in radar polarimetry, the scattering matrix characterizes the target-scattered response in terms of either complex voltage measurements or relative power measurements. For voltage measurements, two coordinate systems are used to specify the incident wave at the transmitter, the scattered wave at the target, and the received wave. The choice of the coordinate system defines the polarization of these wave components, and influences the scattering matrix. In the forward-scatter alignment (FSA) convention, two local right-handed coordinate systems are defined at the transmitter and the target relative to the propagating wave, with the z -axis pointed along the direction of propagation [11]. A third co-ordinate system is defined at the receiver with the z -axis pointed along the line of sight from the target into the receiver. FSA system is employed in bistatic measurements of the scattering matrix, when the transmitter and the receiver antennas are not collocated. In the back-scatter alignment (BSA) coordinate

system, the local coordinate systems at the transmitter and the target are the same as in the FSA convention, but the one at the receiver is defined as a right-handed system with the z -axis pointed away from the receiver toward the target. For a monostatic or backscattering configuration, when the transmitter and the receiver antennas are collocated, both FSA and BSA systems coincide. Typically FSA convention is used to represent the scattering matrix for the bistatic mode, including forward scattering, whereas both BSA and FSA conventions can be employed in the monostatic case, although it is physically intuitive to use the BSA. The Sinclair matrix is used to describe the scattering matrix with a monostatic radar [27]. The Jones matrix is used to describe bistatic scattering matrix [20], [33] (see also Section 6.4 in [3]).

A second approach for scattering matrix theory is based on the description of transmitted and received back-scattered energy in terms of real power measurements, and leads to Kennaugh and Mueller matrices for monostatic and bistatic cases, respectively [14, 34]. The advantage of power measurements is that the elimination of the absolute phase from the target means that the power-related parameters become incoherently additive [11]. Either the voltage-measured Sinclair matrix or the power-measured Kennaugh matrix leads to the same result for completely polarized target returns, which occur when (a) the target geometry as projected on the radar antenna aperture does not change with time (e.g., stationary target) or (b) the target or the intervening medium does not depolarize the scattered signal. In this case, for both formulations, the scattering matrices may be expressed in terms of a common set of five physically significant parameters [14, 19, 23]. However, the Kennaugh matrix is especially useful for partially polarized return signals, which occur when the target geometry varies with time. In this case, the time-averaged Kennaugh matrix is described by nine parameters derived from the Stokes vectors of transmit and receive polarization states [14]. Next, in Section 13.3.2, the five significant parameters of the Sinclair formulation are derived by appealing to radar polarimetry theory based on eigenvector decomposition, introduced by Kennaugh first [13] and later generalized by Huynen [14]. In Section 13.3.3, the Kennaugh matrix is extended to partially polarized target returns, and the similarity between Kennaugh and Mueller matrices is discussed. The theory and notation used to describe the scattering matrix in terms of physically significant parameters follow closely Huynen's seminal exposition on phenomenological theory of radar polarimetry [14].

13.3.1 Sinclair Formulation

The scattering matrix characterizes a radar target's scattered response to an incident wave and describes how the polarization, amplitude and phase of an incident wave are transformed by backscattering from a target. The scattering matrix is a 'transfer function' of the target and hence is independent of the polarization states of the transmit and receive channels of the radar. A monostatic radar transmitting and receiving two orthogonal polarization states is assumed in the following discussion. The target-scattered signal received by a radar can be expressed as [3, 14, 23]

$$V_{rt} = \mathbf{S} \hat{\mathbf{e}}_t \cdot \hat{\mathbf{e}}_r \quad (13.26)$$

where \mathbf{S} is the *Sinclair* scattering matrix, measured in meters, and $\hat{\mathbf{e}}_t$ and $\hat{\mathbf{e}}_r$ are the normalized polarization unit vectors of the transmit and receive channels of the radar. The RCS of the target is given by

$$\sigma_{rt} = |V_{rt}|^2 \quad (13.27)$$

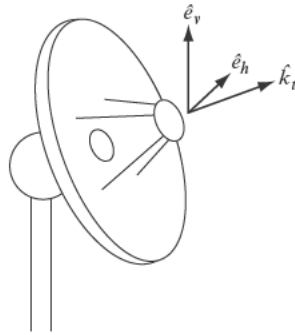


FIGURE 13-6 ■
Illustration of a linear polarization basis oriented in transmitter coordinate system.

The scattering matrix is specified by the basis used to represent the polarization states. The most commonly used basis coordinate system for the measurement of a target scattering matrix is either linear (in terms of orthogonal vertical or horizontal Jones vectors) or circular (right-handed and orthogonal right-handed, or left-handed and orthogonal left-handed Jones vectors). The orthogonal right-handed vector component is similar to left-handed component, except for a phase term [11]. Let \hat{e}_v and \hat{e}_h be orthogonal Jones vectors in the vertical and horizontal directions, respectively, such that $\hat{k}_t = \hat{e}_h \times \hat{e}_v$ defines the direction of wave propagation toward the target, as depicted in Figure 13-6.

Any arbitrary coordinate direction may be chosen for the reference polarization. Thus, choosing x-axis ($\tau = 0$) as the direction of horizontal polarization, one may define the linear basis set in terms of the two Cartesian components

$$\hat{e}_h = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \hat{e}_v = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (13.28)$$

The circularly polarized basis may be represented in terms of the linear basis using unitary transformations as described in [11]. For example, the left-handed orthogonal pair may be written as

$$[\hat{e}_l \quad \hat{e}_{l\perp}] = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix} [\hat{e}_h \quad \hat{e}_v] = T_{LC} [\hat{e}_h \quad \hat{e}_v] \quad (13.29)$$

For the right-handed orthogonal components, j should be changed to $-j$ in the unitary transformation matrix in (13.29). In terms of Cartesian components, the circularly polarized basis is given by

$$\begin{aligned} \hat{e}_l &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ j \end{bmatrix}, \hat{e}_{l\perp} = \frac{1}{\sqrt{2}} \begin{bmatrix} j \\ 1 \end{bmatrix} \\ \hat{e}_r &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -j \end{bmatrix}, \hat{e}_{r\perp} = \frac{1}{\sqrt{2}} \begin{bmatrix} -j \\ 1 \end{bmatrix} \end{aligned} \quad (13.30)$$

It is emphasized that the two orthogonal left-circularly polarized components have been normalized such that $|\hat{e}_l \cdot \hat{e}_l^*| = |\hat{e}_{l\perp} \cdot \hat{e}_{l\perp}^*| = 1$. A similar relation holds for the right-handed sense. It may also be verified that $|\hat{e}_l \cdot \hat{e}_{l\perp}^*| = |\hat{e}_r \cdot \hat{e}_{r\perp}^*| = 0$. The correspondence between (13.20) and (13.30) may be observed by noting that for circular polarization $\tau = \pi/4$ (due to symmetry, orientation angle is undefined). Any arbitrary elliptical polarization can be represented in terms of these two orthonormal bases: the linear basis set (13.28) or

the circular basis set (13.30). If a linear basis is used, the scattering matrix is given by

$$\mathbf{S}_L = \begin{bmatrix} S_{HH} & S_{VH} \\ S_{HV} & S_{VV} \end{bmatrix} \quad (13.31)$$

and a similar representation follows in terms of the circular basis. For a reciprocal medium (e.g., no Faraday rotation of polarization), $S_{VH} = S_{HV}$. It is possible to transform the scattering matrix from the linear basis to the circular basis using

$$\mathbf{S}_C = \mathbf{T}_{LC}^\dagger \mathbf{S}_L \mathbf{T}_{LC} \quad (13.32)$$

where the matrix \mathbf{T}_{LC} is defined in (13.29), and the superscript \dagger denotes conjugate transpose of the matrix. In (13.31), S_{rt} , $r, t = V$ or H , corresponds to the complex scattering amplitude with polarisation t on transmit and polarisation r on receive; for co-polarized channels $r = t$, and for cross-polarized channels $r \neq t$. The absolute phase of the matrix \mathbf{S} does not affect either the wave polarization state or the received signal power. Therefore, one of the elements in \mathbf{S} (say, S_{HH}) may be used as phase reference for the other elements, thus resulting in a target scattering matrix with relative phase

$$\mathbf{S}_{SMR} = \mathbf{S} e^{j\phi_{HH}} \quad (13.33)$$

Without specific reference to (13.33), the relative scattering matrix will be assumed in the remainder of this section.

13.3.2 Optimal Polarizations

The transmit polarization which maximizes or minimizes the total backscattered power can be determined by solving the pseudo-eigenvalue problem [14]

$$\mathbf{S} \hat{\mathbf{e}}_i = \lambda_i \hat{\mathbf{e}}_i^*, i = 1, 2 \quad (13.34)$$

where $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ are two distinct orthogonal eigenvectors (unit magnitude) associated with complex eigenvalues λ_1 and λ_2 , respectively. It can be shown that the eigenvectors corresponding to orthogonal elliptically polarized states are given by (see (13.20)) [11]

$$\begin{bmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 \end{bmatrix} = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \begin{bmatrix} \cos \tau & j \sin \tau \\ j \sin \tau & \cos \tau \end{bmatrix} \begin{bmatrix} \hat{\mathbf{e}}_h & \hat{\mathbf{e}}_v \end{bmatrix} \triangleq \begin{bmatrix} m & n \\ -n^* & m^* \end{bmatrix} \begin{bmatrix} \hat{\mathbf{e}}_h & \hat{\mathbf{e}}_v \end{bmatrix} \quad (13.35)$$

The matrix in (13.35) transforms the linear polarization basis into an elliptical polarization basis. It is evident from the second equality that $\hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_2^* = 0$, $|\hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_1^*| = 1$, and $|\hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_2^*| = 1$. Also, $\hat{\mathbf{e}}_2(\psi, \tau) = \hat{\mathbf{e}}_1(\psi + \pi/2, -\tau)$. It then follows that the matrix $\mathbf{U} = [\hat{\mathbf{e}}_1 \quad \hat{\mathbf{e}}_2]$ is unitary, that is, $\mathbf{U}^T \mathbf{U}^* = \mathbf{I}$, where \mathbf{I} is the identity matrix, and superscript T denotes the transpose. The unitary transformation $\mathbf{U}^T \mathbf{S} \mathbf{U}$ diagonalizes the scattering matrix, resulting in

$$\mathbf{U}^T \mathbf{S} \mathbf{U} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \triangleq \mathbf{S}_d \quad (13.36)$$

Inverting (13.36) using the property $\mathbf{U}^{-1} = \mathbf{U}^\dagger$, where \mathbf{U}^\dagger denotes the Hermitian adjoint or conjugate transpose, one obtains the scattering matrix from the eigendecomposition

in (13.34) as

$$S = U^* S_d U^\dagger \quad (13.37)$$

Given a transmit polarization state, the scattering matrix elements for co-polarized and cross-polarized returns are obtained from the target geometry and the return signal. Knowing S , one may calculate its eigenvalues and the eigenvectors and determine the transmit polarization state that produces the maximum return [13, 14].

Let the transmit polarization state that yields the maximum backscattered power from the target be denoted in terms of the two eigenvectors in (13.35) as $\mathbf{e}_t = a_1 \hat{\mathbf{e}}_1 + a_2 \hat{\mathbf{e}}_2$. The total scattered power is given by

$$W_s = \mathbf{E}_s \cdot \mathbf{E}_s^* = \mathbf{S} \mathbf{e}_t \cdot (\mathbf{S} \mathbf{e}_t)^* \quad (13.38)$$

Using (13.34) and the orthogonality of the two eigenvectors, it follows that

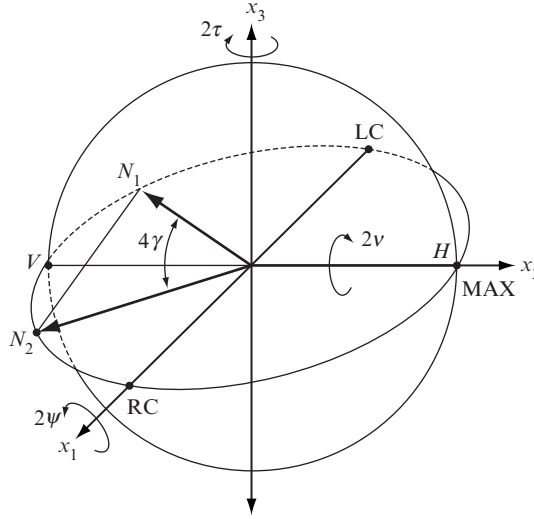
$$W_s = |a_1|^2 |\lambda_1|^2 + |a_2|^2 |\lambda_2|^2 = (|a_1|^2 + |a_2|^2) |\lambda_1|^2 - |a_2|^2 (|\lambda_1|^2 - |\lambda_2|^2) \quad (13.39)$$

Requiring without loss of generality that $|\lambda_1|^2 > |\lambda_2|^2$, (13.39) states that the backscattered power is maximum when $a_2 = 0$. Therefore, the eigenvector $\hat{\mathbf{e}}_1$ in (13.35) corresponds to the transmit polarization state $\hat{\mathbf{e}}_1(\psi, \tau)$ that maximizes the backscattered power. If the receive antenna has the same polarization state, then the total received power would be maximum as well (see (13.26)), hence, $\hat{\mathbf{e}}_1(\psi, \tau)$ is called Copol max [13, 35–37]. The angles ψ and τ denote the target orientation angle and the ellipticity angle, respectively, for maximum power. The orthogonal polarization state $\hat{\mathbf{e}}_2 = \hat{\mathbf{e}}_1(\psi + \pi/2, -\tau)$ results in a secondary maximum corresponding to a saddle point of the scattering matrix, denoted as Copol saddle [36, 37]. Equivalently, these states also correspond to two cross-polarization nulls and are referred to as null polarizations [14, 23]. It can be shown that the corresponding eigenvalues are given by [14]

$$\begin{aligned} \lambda_1 &= m e^{j2v} \\ \lambda_2 &= m \tan^2 \gamma e^{-j2v} \end{aligned} \quad (13.40)$$

where m is the maximum amplitude of the received signal and its square indicates an overall measure of target size or RCS, v is called the *target skip angle* and is related to the number of bounces of the signal reflected off the target ($-\pi/4 \leq v \leq \pi/4$), and γ is called the target polarizability angle ($0 \leq \gamma \leq \pi/4$), indicating the target's ability to polarize incident unpolarized radiation (0° for fully polarized return and 45° for unpolarized return). The angle τ determines the ellipticity of the optimal polarization eigenvector $\hat{\mathbf{e}}_1$ or $\hat{\mathbf{e}}_2$, and is a measure of target symmetry with respect to right- and left-circular polarizations (0° for symmetric and 45° for totally asymmetric). It is also referred to as target helicity angle [14]. Likewise, ψ is called the orientation angle and is a measure of the orientation of the target around the radar line of sight. For roll-symmetric targets, such as a cone, if the axis of symmetry can be aligned to be coplanar with a ray from radar to target, the target coordinate system may be chosen such that the scattering matrix becomes diagonal (no cross-polarization terms), and the orientation angle can be determined directly from the polarization eigenvectors [14]. These five parameters (m, ψ, v, τ, γ) uniquely define the scattering matrix (within a constant phase factor; see (13.33)), and provide significant physical insight into the target scattering process [13, 14].

FIGURE 13-7 ■
Polarization fork.
(From Huynen [14].
With permission.)



The set of null polarizations associated with the relative scattering matrix can be represented graphically on the Poincaré sphere. For the moment, assume that the two cross-polarization nulls correspond to orthogonal polarizations $\hat{e}_1 = \hat{e}_h$, $\hat{e}_2 = \hat{e}_v$, which are located on the equator at H and V , respectively (see Figure 13-7). These null polarizations are eigenvectors of S and satisfy the relations $\hat{e}_1 \cdot S\hat{e}_2 = \hat{e}_2 \cdot S\hat{e}_1 = 0$. The co-polarization nulls N_1 and N_2 are located symmetrically with respect to the cross-polarization null V , such that the diameter HV bisects the angle 4γ subtended at the center by the arc N_1VN_2 . The co-polarization null polarizations \hat{n}_1 and \hat{n}_2 satisfy the relations $\hat{n}_1 \cdot S\hat{n}_1 = \hat{n}_2 \cdot S\hat{n}_2 = 0$ [14]. The line segments for the three independent null polarizations at H , N_1 and N_2 are rigidly connected at the center, forming a fork known as Huynen fork or polarization fork [14]. The relative location of the optimal polarizations is invariant with respect to the polarization basis, so that Huynen's fork is polarization-invariant. As the optimal transmit or receive polarization state changes, the scattering matrix undergoes three sequential rotations about the x_1 , x_2 , and x_3 axes by angles 2ψ , 2ν , and 2τ , respectively. Mathematically, these rotations are described by [23] (see (13.37))

$$S = U^*(\psi, \tau, \nu) \begin{bmatrix} m & 0 \\ 0 & m \tan^2 \gamma \end{bmatrix} U^\dagger(\psi, \tau, \nu) \quad (13.41)$$

Using Pauli spin matrices J , K , and L , one may write

$$U(\psi, \tau, \nu) = e^{\psi J} e^{\tau K} e^{\nu L} \quad (13.42)$$

$$J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, K = \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix}, L = \begin{bmatrix} -j & 0 \\ 0 & j \end{bmatrix}$$

Therefore

$$e^{\psi J} = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix}, e^{\tau K} = \begin{bmatrix} \cos \tau & j \sin \tau \\ j \sin \tau & \cos \tau \end{bmatrix}, \text{ and } e^{\nu L} = \begin{bmatrix} e^{-j\nu} & 0 \\ 0 & e^{j\nu} \end{bmatrix} \quad (13.43)$$

After these three rotations of the fork inside the Poincaré sphere, the intersection of the handle of the fork with the sphere represents the eigenvector \hat{e}_1 of S , and the null

polarizations $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ correspond to intersection of the two tines of the fork with the sphere at the points N_1 and N_2 .

13.3.3 Partial Polarization and Mueller Matrix

An arbitrary polarization state $\hat{\mathbf{h}}(\tau, \psi)$ is given by (13.20), where the geometrical parameters τ and ψ of the polarization ellipse do not vary with time for a fully polarized wave. Thus, the polarization state remains constant over time. A wave whose polarization state varies with time in a random manner is said to be partially polarized. Partial polarization also arises due to spatial variation, specifically in radar imaging of distributed targets that are homogeneous but have limited or no correlation from pixel to pixel. Due to time-varying fluctuations of both the propagation medium and the target, the wave scattered by the target may be partially polarized even though the wave transmitted by the radar may be fully polarized. Considering time variation, the polarization state of the scattered wave may be written as

$$\hat{\mathbf{h}}(\tau, \psi) = \begin{bmatrix} h_x(t) \\ h_y(t) \end{bmatrix}, \tau = \tau(t), \psi = \psi(t) \quad (13.44)$$

where $h_x(t)$ and $h_y(t)$ are complex envelopes of the orthogonal field components, and the time dependence of tilt and ellipticity angles is explicitly stated. The time variations of the scattered wave can be attributed to random processes. It is usual to assume that $h_x(t)$ and $h_y(t)$ are zero-mean complex random variables, with phases uniformly distributed and independent of moduli [14]. This assumption agrees with the hypothesis that the scattered wave is given by the incoherent sum of many waves with different polarizations. Therefore, the random time variation of the polarization state can be described by correlations between the field components.

A partially polarized plane wave can be described through the time or ensemble average of the Stokes vector in (13.21) to obtain

$$\mathbf{g} = \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} \langle |E_x|^2 \rangle + \langle |E_y|^2 \rangle \\ \langle |E_x|^2 \rangle - \langle |E_y|^2 \rangle \\ \langle 2 \operatorname{Re}(E_x E_y^*) \rangle \\ \langle 2 \operatorname{Im}(E_x E_y^*) \rangle \end{bmatrix} = \begin{bmatrix} \langle g_0 \rangle \\ \langle g_0 \cos 2\tau \cos 2\psi \rangle \\ \langle g_0 \cos 2\tau \sin 2\psi \rangle \\ \langle g_0 \sin 2\tau \rangle \end{bmatrix} \quad (13.45)$$

where $\langle ab^* \rangle$ denotes ensemble average of the correlation between field components a and b . The total average power, given by g_0 , satisfies the condition [34]

$$g_0 \geq \sqrt{g_1^2 + g_2^2 + g_3^2} \quad (13.46)$$

where the equality holds when the wave is fully polarized, the inequality ($>$ sign) is valid for a partially polarized wave, and $g_1 = g_2 = g_3 = 0$ for an unpolarized wave.

If any of the parameters g_1, g_2 , or g_3 has a nonzero value, it indicates the presence of a polarized component in the wave. The degree of polarization is defined as the ratio of the completely polarized power to the total average power and is given by

$$d = \frac{\sqrt{g_1^2 + g_2^2 + g_3^2}}{g_0}, \quad 0 \leq d \leq 1 \quad (13.47)$$

A partially polarized wave may be regarded as the sum of a completely polarized wave and a completely unpolarized wave. Thus, the Stokes vector of a partially polarized wave may be written in normalized form as

$$\bar{\mathbf{g}} = \begin{bmatrix} 1 \\ g_1/g_0 \\ g_2/g_0 \\ g_3/g_0 \end{bmatrix} = \begin{bmatrix} d \\ d \cos 2\tau \cos 2\psi \\ d \cos 2\tau \sin 2\psi \\ d \sin 2\tau \end{bmatrix} + \begin{bmatrix} 1-d \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (13.48)$$

where the first vector on the right represents the polarized part, and the second term represents the unpolarized part of the wave. Equation (13.48) forms the basis of polarization decomposition theorems that allow one to write the partially polarized scattered signal in terms of a completely polarized part corresponding to the target structure and a residue or noise term (representing, e.g., clutter, motion) [13, 14, 23].

Let \mathbf{g} be the Stokes vector corresponding to the transmit antenna polarization state $\hat{\mathbf{e}}_t$, and \mathbf{h} be the Stokes vector corresponding to the receive antenna polarization state $\hat{\mathbf{e}}_r$. From (13.26) and (13.27), it is evident that the backscattered power received by the radar is proportional to RCS of the target and may be written as

$$P = |V_{rt}|^2 = |S(m, \psi, v, \tau, \gamma) \hat{\mathbf{e}}_t(\psi_a, \tau_a) \cdot \hat{\mathbf{e}}_r(\psi_b, \tau_b)|^2 \quad (13.49)$$

where the argument of the scattering matrix consists of the five significant Huynen parameters describing the scattered wave from the target (see Section 13.3.2), (ψ_a, τ_a) and (ψ_b, τ_b) denote the orientation and ellipticity angles, respectively, of the transmit and receive antenna polarization states. A completely polarized wave is assumed for the moment. Huynen [14] showed that (13.49) may be concisely written in terms of the Stokes vectors of the two antenna polarizations, as defined in (13.21), and the Kennaugh matrix defined by the target parameters, as

$$P = \mathbf{K} \mathbf{g} \cdot \mathbf{h} \quad (13.50)$$

The 4×4 real symmetric Kennaugh matrix represents power measurements and can be expressed in terms of the five target parameters, m, ψ, v, τ, γ , as [14, 23]

$$\mathbf{K} = \begin{bmatrix} A_0 + B_0 & C_\psi & H_\psi & F \\ C_\psi & A_0 + B_\psi & E_\psi & G_\psi \\ H_\psi & E_\psi & A_0 - B_\psi & D_\psi \\ F & G_\psi & D_\psi & -A_0 + B_0 \end{bmatrix} \quad (13.51)$$

where

$$\begin{aligned} A_0 &= Qf \cos^2 2\tau \\ B_0 &= Q(1 + \cos^2 2\gamma - f \cos^2 2\tau) \\ B_\psi &= B \cos 4\psi - E \sin 4\psi \\ C_\psi &= C \cos 2\psi \\ D_\psi &= G \sin 2\psi + D \cos 2\psi \\ E_\psi &= B \sin 4\psi + E \cos 4\psi \\ F &= 2Q \cos 2\gamma \sin 2\tau \end{aligned}$$

$$\begin{aligned}
G_\psi &= G \cos 2\psi - D \sin 2\psi \\
H_\psi &= C \sin 2\psi \\
B &= Q (1 + \cos^2 2\gamma - f(1 + \sin^2 2\tau)) \\
C &= 2Q \cos 2\gamma \cos 2\tau \\
D &= Q \sin^2 2\gamma \sin 4v \cos 2\tau \\
E &= -Q \sin^2 2\gamma \sin 4v \sin 2\tau \\
G &= Qf \sin 4\tau \\
Q &= m^2 / (8 \cos^4 \gamma) \\
f &= 1 - \sin^2 2\gamma \sin^2 2v
\end{aligned} \tag{13.52}$$

For time-invariant targets, the first nine parameters of the Kennaugh matrix (also known as Huynen parameters [23]) are not all independent but are related by the four auxiliary equations

$$\begin{aligned}
2A_0(B_0 + B_\psi) &= C_\psi^2 + D_\psi^2 \\
2A_0(B_0 - B_\psi) &= G_\psi^2 + H_\psi^2 \\
2A_0E_\psi &= C_\psi H_\psi - D_\psi G_\psi \\
2A_0F &= C_\psi G_\psi + D_\psi H_\psi
\end{aligned} \tag{13.53}$$

The parameters $A_0, C_\psi, G_\psi, H_\psi$, are fixed and the parameters B_0, B_ψ, E_ψ, F , are calculated using (13.53), thereby reducing the nine parameters from (13.52) into the five desired ones, m, ψ, v, τ, γ .

It is interesting to note a subtle difference between the Kennaugh matrix in (13.51) and the equivalent Huynen's formulation of the Mueller matrix in [14]. As described in [38], instead of the conventional Stokes vector defined in (13.45), if one uses the representation

$$\mathbf{s} = \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} \langle |E_x|^2 \rangle + \langle |E_y|^2 \rangle \\ \langle 2 \operatorname{Im}(E_x^* E_y) \rangle \\ \langle |E_x|^2 \rangle - \langle |E_y|^2 \rangle \\ \langle 2 \operatorname{Re}(E_x^* E_y) \rangle \end{bmatrix} = \begin{bmatrix} \langle g_0 \rangle \\ \langle g_0 \sin 2\tau \rangle \\ \langle g_0 \cos 2\tau \cos 2\psi \rangle \\ \langle g_0 \cos 2\tau \sin 2\psi \rangle \end{bmatrix} \tag{13.54}$$

the Kennaugh matrix in (13.51) must be rearranged to yield the Mueller matrix given by

$$\mathbf{M} = \begin{bmatrix} A_0 + B_0 & F & C_\psi & H_\psi \\ F & -A_0 + B_0 & G_\psi & D_\psi \\ C_\psi & G_\psi & A_0 + B_\psi & E_\psi \\ H_\psi & D_\psi & E_\psi & A_0 - B_\psi \end{bmatrix} \tag{13.55}$$

Huynen developed the phenomenological theory of radar targets based on the Mueller matrix in (13.55). This difference between Kennaugh and Mueller matrices simply involves a reordering of elements in the Stokes vector. The Stokes vector defined in (13.45) denotes the standard expression used in the literature, and is adopted throughout this chapter. As stated earlier, for fully polarized returns, the two matrix representations yield the same scattering parameters.

With time-variant targets, the nine Huynen target parameters in (13.52) become independent if one considers the time-averaged Mueller matrix

$$\mathbf{R} = \langle \mathbf{M}(t) \rangle \quad (13.56)$$

Analogous to (13.48), Huynen showed that the matrix \mathbf{R} may be decomposed in terms of the target and residue parts as

$$\mathbf{R} = \mathbf{M}_0 + \mathbf{N} \quad (13.57)$$

where \mathbf{M}_0 is the symmetric Mueller or Kennaugh matrix defined in (13.51) and (13.53) for a coherently scattering (fixed) target in terms of the five parameters, m, ψ, v, τ, γ , and \mathbf{N} is the residue matrix, also referred to as nonsymmetric noise target [14], which describes target fluctuations. It is emphasized that the target signature fluctuations around a mean stationary target return need not be unpolarized as in (13.48) and are independent of transmit and receive polarization states. To define the matrix \mathbf{N} , one writes the time-averaged parameters $\langle B_0 \rangle, \langle B_\psi \rangle, \langle E_\psi \rangle, \langle F \rangle$, as

$$\begin{aligned} \langle B_0 \rangle &= B_0^T + B_0^N \\ \langle B_\psi \rangle &= B_\psi^T + B_\psi^N \\ \langle E_\psi \rangle &= E_\psi^T + E_\psi^N \\ \langle F \rangle &= F^T + F^N \end{aligned} \quad (13.58)$$

where $B_0^T, B_\psi^T, E_\psi^T, F^T$, satisfy (13.53). The residue matrix \mathbf{N} is then given by [14]

$$\mathbf{N} = \begin{bmatrix} B_0^N & F^N & 0 & 0 \\ F^N & B_0^N & 0 & 0 \\ 0 & 0 & B^N & E^N \\ 0 & 0 & E^N & -B^N \end{bmatrix} \quad (13.59)$$

It is instructive to discuss the physical significance of Huynen's decomposition theorem in (13.57). Five Stokes parameters define \mathbf{M}_0 , and four Stokes parameters define \mathbf{N} , which totally define \mathbf{R} . This decomposition of a time-varying target into a mean-stationary target described by \mathbf{M}_0 , and a noise target represented by \mathbf{N} , signifies fluctuations of the time-varying target about its mean stationary representation. The time dependence of the target can be due to, for example, target or radar motion or clutter. Therefore, the decomposition theorem is very effective in distinguishing between target-specific attributes of the backscattered return and those induced by motional perturbations and clutter. The amount of motion variance attributed to a target may provide some insight into the nature of the target. For example, in radar images of the target return for a reentry vehicle [39], it has been observed by range-Doppler correlation that the effect of motion on target-specific scattering properties, such as the specular return, can be separated from corresponding effects on clutter. Mathematically, these differences in variations of Doppler between coherent target returns and clutter can be analyzed by examining the characteristics of their respective \mathbf{N} matrices.

Huynen's decomposition has problems of uniqueness and is not used widely within the radar imaging community. Several other orthogonal target decomposition theorems facilitate extraction of the mean stationary target from an average Mueller matrix [40]. The reader is referred to Lee and Pottier [11] for an extensive review of such theorems tested on

synthetic aperture radar (SAR) polarimetric imaging data. Several authors have enhanced Huynen's decomposition and demonstrated effective separation of target return from clutter [2, 5, 11, 18, 38, 40, 41]. Barnes [42] and Holm [43] developed a target decomposition approach in which the covariance or density matrix representing a partially polarized return from the time-varying target is formulated as a sum of three components: (a) a mean target component (matrix) for which the return is completely polarized; (b) a component representing a totally unpolarized return; and (c) a component representing a partially polarized return. The Barnes and Holm decomposition has been shown to yield clutter discrimination comparable in accuracy to that of the Huynen approach [11]. Cloude and Pottier [40] discussed a decomposition in which a 4×4 averaged coherency matrix for noise target has been diagonalized by calculating its eigenvalues and eigenvectors. The diagonal form of the coherency matrix has been shown to result in statistically independent target vectors wherein the effect of noise (or target fluctuations) is removed. The reader is referred to examples in [11] for a discussion of the clutter discrimination potential of this approach.

13.4 | RADAR APPLICATIONS OF POLARIMETRY

Target detection, discrimination, and recognition pertain to the process by which the radar separates and classifies the return from a resolution cell as the response due to target (object of interest), clutter (objects of lesser interest), or noise (see Chapter 9 in [3]). In the detection process, well-known techniques for thresholding noise and clutter amplitude or exploitation of Doppler frequency are used to effect an initial separation of target and clutter returns [4]. After detecting a target, the radar attempts to discriminate between potential targets of interest and strong target-like responses of clutter, both of which might have filtered through as targets during the detection phase. Frequency agility, Doppler, and polarimetric processing have been used in the discrimination phase. Target recognition pertains to determination whether the detected target belongs to a class of objects that the radar operator is looking for. For example, ballistic missile reentry vehicles and decoys form a class of potential radar objects of interest for detection, while discrimination between the two objects in the presence of clutter belongs to the second phase, and the classification of the missile using features such as nose-tip scattering and material coatings belongs to the recognition phase. In this section, essential polarization behavior of different radar objects such as targets, land and sea clutter are described.

13.4.1 Targets

High-frequency backscattering from complex targets is essentially a local phenomenon, such as specular scattering or edge diffraction. In fact, mathematical analyses and measurements show that the backscattered wave can be essentially regarded as the coherent sum of contributions arising from a finite number of discrete scattering centers [44]. The wave scattered by each such point source depends on the shape and constitutive material properties of a small surface around the scattering center. For example, scattering centers can be associated with specular reflection off a large convex surface and diffraction arising from discontinuities (e.g., base edge of a cone) and changes in curvature of the body. The position of the backscattering centers is sensitive to target aspect, especially in the case of specular reflection. Relative motion of the scattering centers also affects the phase of the constituent waves in the coherent summation.

The wave propagating from a scattering center generally does not have the same polarization as the incident wave. In general, targets possess an asymmetrical shape and a partially conducting surface; therefore, depolarization can occur. In other words, assuming a completely polarized incident wave, the degree and type of polarization of the scattered wave can both be different because of variation in Stokes parameters induced by changes in local radii of curvature and material properties of the scattering surface. For example, the wave reflected by an ellipsoid will have a different tilt angle from the incident wave because of the differences in local radii curvature in the two orthogonal directions at the specular reflection point.

Giuli [2] presented radar polarimetric data on an aircraft target tracked from takeoff until it gained altitude of approximately 30 km (also see Chapter 9 in [3]). The monostatic scattering matrix was measured in S band with a pulse length of $1\ \mu\text{s}$, a pulse repetition frequency of 1 kHz, and a dwell time on target of about 20 ms. Right circular polarization was transmitted and both senses of CP were received. The polarization state of each received pulse was plotted on an orthographic polarization chart for two track positions of the aircraft. The first one corresponds to receive polarizations states during one dwell time (20 pulses in 20 ms) in which the aircraft was moving approximately transverse to the radar line of sight. These points are tightly grouped, with maximum deviation in orientation angle of about 10° and that in ellipticity angle around 5° . The second track was scanned 20 s after the previous one, with the plane turning away from the transverse direction. The maximum deviations of the tilt and ellipticity angles in this group were considerably larger than the previous track, corresponding to 35° and 20° , respectively. Thus, the target echo polarization states did not change much during the dwell time of any one scan but changed considerably between scans, indicating a small degree of depolarization with time.

13.4.1.1 Target Feature Extraction

In the target classification phase, the shape and type of the target can be determined by extracting structural features of the object, such as seams and sharp edges, from the radar return using high-resolution radar imaging [45]. Two-dimensional spectral analysis applied to a sequence of wideband radar returns creates a range/range-rate image of the target, pixelated in terms of radar resolution cells [24–26]. The two dimensions express the variation in time, over the pulse sequence, of the frequency components in each return separately. Typically each 2-D spectral component in the data is associated with a particular scatterer on the target. The phase behavior, as a function of frequency and time, determines the scatterer's range and range rate. The amplitude coefficient determines its radar cross section.

Conventionally, the 2-D Fourier transform is used for this spectral analysis. In recent years, more powerful techniques within the realm of super-resolution methods [46], such as Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [24, 47], subspace techniques such as the state space method (SSM) [26, 48], have emerged that achieve a significant improvement in spectral resolution. Another important advantage of these techniques is that they produce no range or range-rate sidelobes to confuse image interpretation, thereby enabling accurate feature recognition using robust estimates of the amplitude and phase of the individual scatterers.

Fully polarimetric monostatic RCS (i.e., magnitude and phase) of both metallic (aluminum) and dielectric-coated cones with spherical nose tips, illustrated in Figure 13-8, has been measured in a static range at Lincoln Laboratory [24–26, 39]. The metallic cone

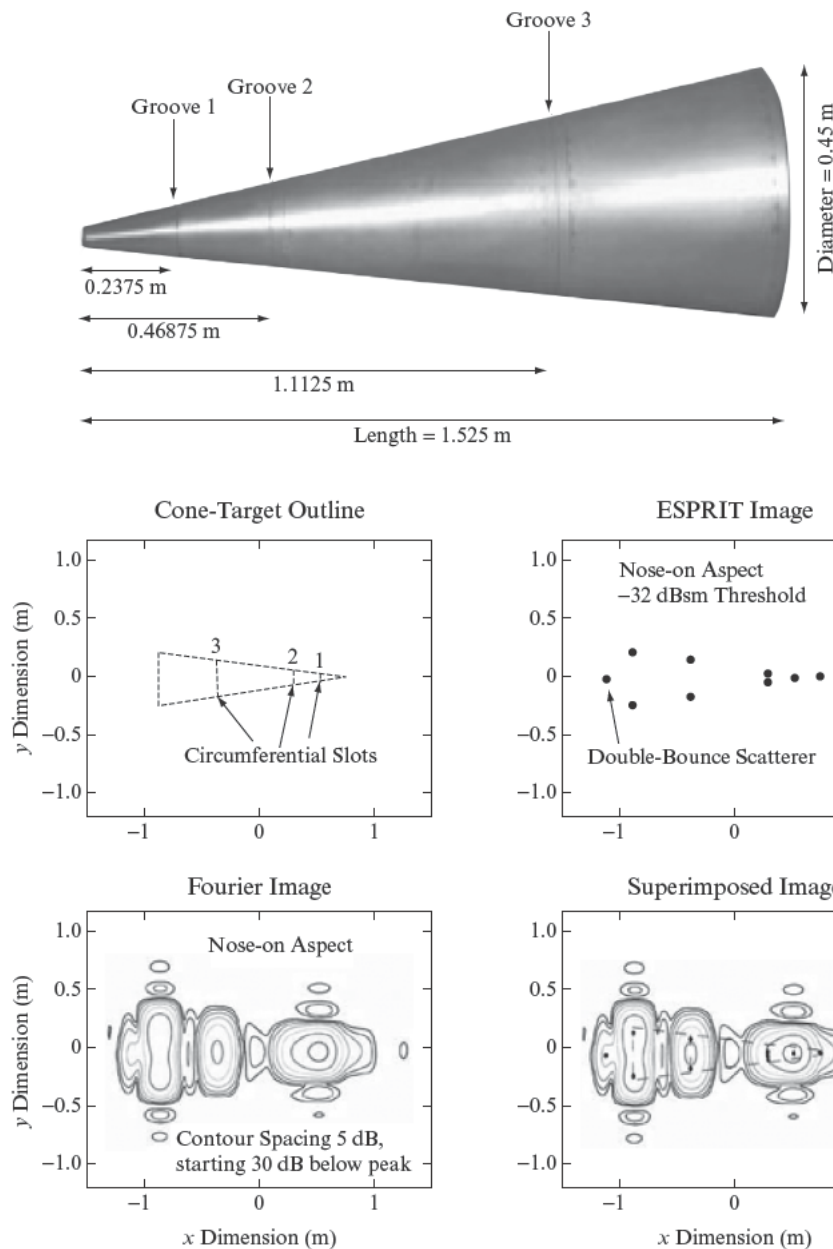


FIGURE 13-8 ■ Metallic conical test target with three machined grooves. (From Burrows [24], Naishadham and Piou [26], and Cuomo et al. [39]. With permission.)

FIGURE 13-9 ■ Fourier and ESPRIT X band radar images of an aluminum round-nosed cone target at nose-on aspect derived from measured static-range data of 1 GHz bandwidth. The target's rotation axis and the radar's magnetic-field polarization on transmit and receive were vertical (HH polarization); the target's body axis of symmetry and the radar line of sight were horizontal. (From Burrows [24]. With permission.)

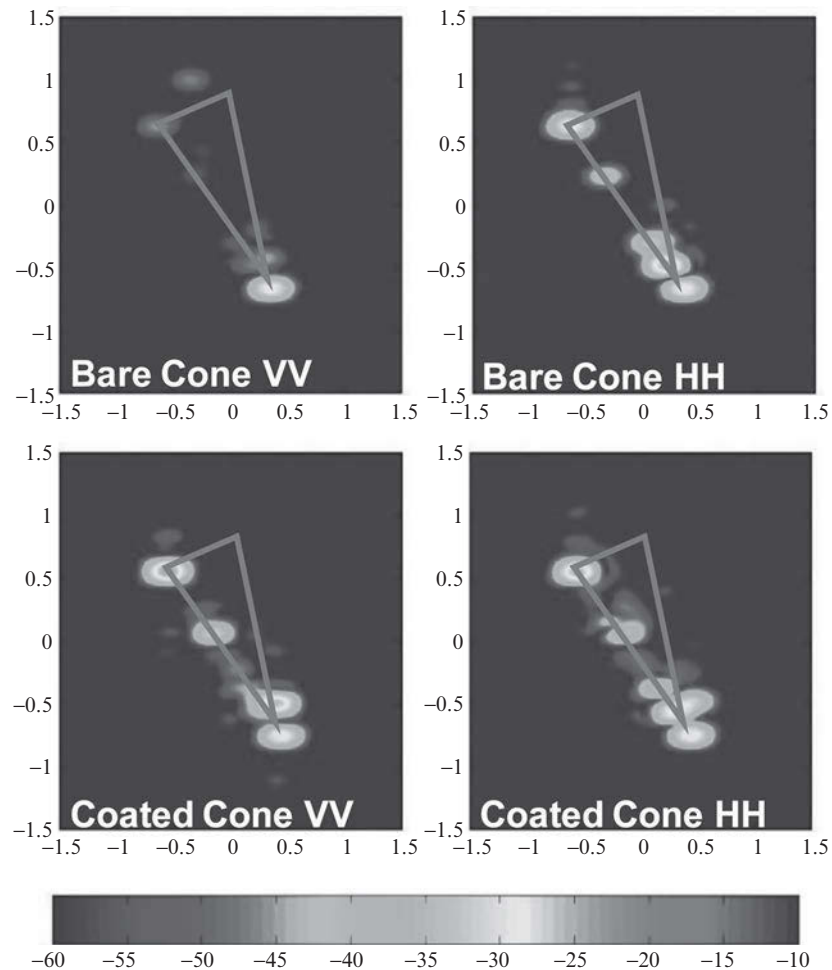
is 1.6 m long and has a base diameter of 0.45 m. It is machined from four frusta (of the same cone angle) joined together at locations marked by three rectangular grooves or slots. The first groove is 3.13 mm deep and 6.25 mm wide, while the second and third grooves measure 1.88 mm \times 3.75 mm and 1.25 mm \times 2.5 mm, respectively. The coated cone, with a core of identical dimensions to the metallic one, contains an outer dielectric layer of 4.57 mm thickness and dielectric constant, $\epsilon_r = 3.55 - j0.55$.

Figure 13-9 compares the X band images of the conical target extracted by ESPRIT and by Fourier processing from the same block of static-range RCS data, covering a narrow angular sector centered on nose-on aspect, with data bandwidth of 1 GHz [24]. The

ESPRIT image simply consists of the estimated locations of the principal scatterers. The Fourier image, since these locations are not directly available, is the conventional contour plot of the discrete 2-D Fourier transform of the windowed radar data. The improvement that ESPRIT processing provides, both in sharpness and spectral splatter, is striking. It images accurately all three slots and resolves reliably the two edges of slot 3, the largest in diameter. The phantom scatterer lying behind the base of the cone is the result of the doubly diffracted ray that crosses the flat base of the target from one edge to the other and returns to the radar along the line of sight [49]. These doubly diffracted rays cause extended returns in radar target signatures because the corresponding scattering centers are located beyond the target [26].

As an example on how polarization-dependent scattering from lossy dielectric materials, such as heat-resistant coatings on reentry vehicles, can be used to discriminate between targets of interest and fake targets or decoys, polarimetric behavior of the conical test target (Figure 13-8) is considered for the case when it coated with a dielectric material. Figure 13-10 compares the X band (10 GHz) monostatic RCS images of both bare and coated targets as a function of transmit and receive polarization states. Only the co-polarized states HH and VV are depicted.

FIGURE 13-10 ■ Radar images (range/cross-range) of co-polarized returns for bare and coated cones.



The radar line-of-sight aspect is 30° relative to the axis of symmetry of the cone. The three grooves, base edge and the nose tip can be identified from the images using the cone geometry depicted in Figure 13-8. It is evident that bare targets tend to polarize the target return, because there is clear distinction between the VV and HH returns at edges and joints. Except for the strong nose-tip return, which is the closest in range to the radar, all other features, including the three grooves, are masked in the VV return. In contrast, material coating tends to depolarize the target return, and the distinction between VV and HH states is less clear, particularly at the higher frequencies. The contrast between bare and coated cones is larger for VV than HH, which makes a strong case for using polarimetric behavior of the target return as a reliable means for feature identification and target discrimination.

Based on the images presented in Figure 13-10, it is clear that depolarization reduces the actionable information that can be gained from polarimetric probing at the aspect angle measured. Are there other aspect angles and target scattering features that one can exploit to provide better contrast between coated and bare targets? In target identification, mostly features such as specular returns, base edges, and joints, which provide visible change in signature at few specific aspect angles, are employed in the discrimination phase. However, these features also produce diffused scattering, which makes the corresponding scattering centers highly dependent on the aspect angle. By extensive investigation of the physics of scattering from coated targets [26, 50, 51], it has been determined that unconventional scattering mechanisms, such as creeping waves and multiply diffracted waves, produce less depolarization than base edges.

As an example, the polarimetric behavior of a dielectric coated cylinder, 0.85 m long and 0.1213 m in radius, is considered next. The dielectric coating has the same parameters as those for the conical target (Figure 13-8). Figure 13-11 depicts the main scattering mechanisms of specular reflection and creeping waves, both of which are strongest at normal incidence.

The radar images at normal incidence and a frequency of 4 GHz are shown in Figure 13-12, comparing VV and HH polarized returns for the dielectric-coated cylinder. Interestingly, while there is no noticeable difference in specular scattering, the creeping wave, occurring at a range of 0.25 m, is not present in the HH return. This should be contrasted with Figure 13-10, where there is no significant difference between VV and HH returns for the coated conical target.

By lowering the amplitude threshold, it has been shown [51] that there exists a doubly diffracted wave at the cylinder base edge for normal incidence, which is stronger for

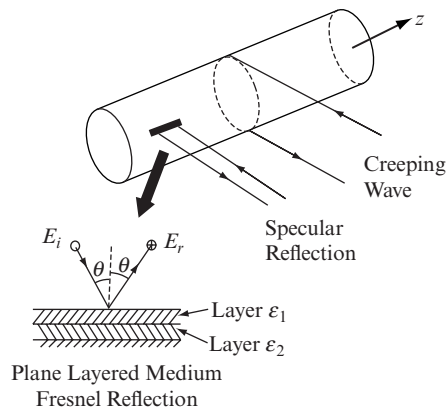
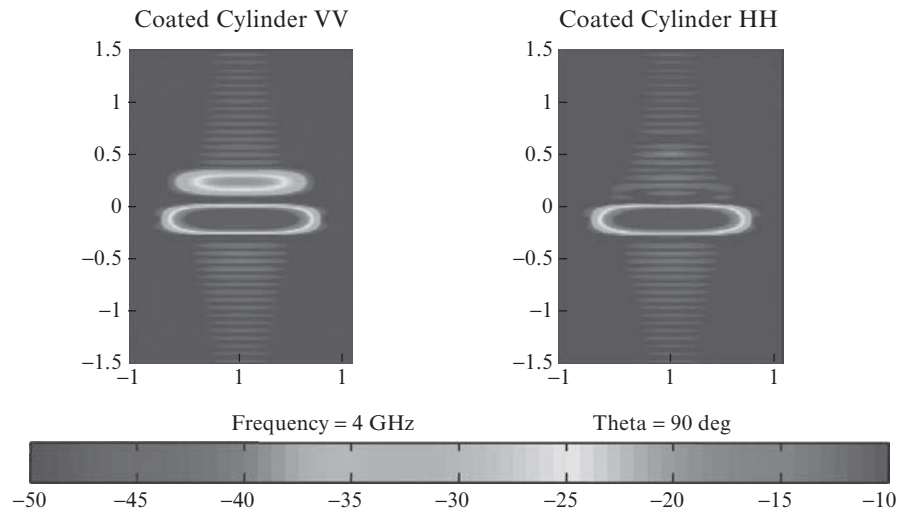


FIGURE 13-11 ■ Specular reflection, based on physical optics, and creeping wave paths for a cylinder. (From Naishadham and Piou [51]. With permission.)

FIGURE 13-12 ■ Radar images (range/cross-range) of co-polarized returns for coated cylinder. (From Naishadham and Piou [51]. With permission.)



HH polarization than VV. Therefore, variation in the target return with aspect angle, sensed as a function of the transmit and receive polarization states, provides interesting physics-based information that can be applied to discriminate between coated and uncoated targets.

13.4.2 Rain Clutter

Much of the clutter surrounding a target is due to objects that move relative to each other, such as ocean waves, rainfall, and foliage movement caused by wind. Clutter can appear within the resolution cell of the target; if it is not separated from the target return, clutter can cause missed detections and false alarms. The scattering matrix for clutter is time varying in general, but for typical radar observation times (few ms) the matrix may be assumed constant. Rain does not depolarize the incident wave within time scales comparable to the observation time. Precipitation clutter is known to be very sensitive to antenna polarization, and the choice of circular polarization as a means of reducing rain clutter by as much as 30 dB is widely recognized [2]. The rationale behind the use of circular polarization is that the total return from a spatial resolution cell, within the precipitation volume, is produced by a large number of spherically shaped hydrometeors. In theory, if a circularly polarized wave is incident on either a conducting or dielectric sphere, the backscattered wave has the opposite rotation sense, and polarization cancellation may be used at the receiver to reduce the clutter from raindrops. In practice, however, the effectiveness of such cancelation of precipitation clutter is limited, ranging approximately from 10 to 35 dB, since the assumption of spherical raindrops is only approximately met. Larger raindrops, approximating oblate spheroids, occur for increasing rainfall rates, and in the presence of wind the shape may become quite asymmetrical and complex [3]. Multipath propagation due to reflection off land or water (e.g., sea, river, lake) surface can cause significant depolarization of the incident wave for heavy rainfall, especially in windy conditions. Except for light rain, a significant improvement of clutter attenuation can be achieved by adaptation of antenna polarization. An improvement in clutter cancelation by about 12 dB has been reported for heavy rain using elliptical polarization (compared with circular) [52]. However, considering the additional complexity of radar hardware and

the waveform processing, the advantages of using circular polarization far outweigh its limitation in clutter attenuation for heavy rain. Therefore, the use of circular polarization to reduce rain clutter is still widespread. The reader is referred to [53, 54] for a comprehensive treatment of radar meteorology and accurate polarimetric measurements of the rain clutter.

13.4.3 Ground Clutter

Ground clutter due to backscattering from land or sea frequently limits radar target visibility. It usually introduces depolarization as a consequence of ground surface roughness. Ground clutter is, in general, distributed with a high degree of nonstationarity both in time and space. This indicates considerable depolarization of the backscattered wave. A higher wind velocity and frequency can also significantly increase depolarization and reduce the degree of polarization of distributed ground clutter. For strong point clutter in an urban environment, due to spatial regularity of the ground surface or the prevailing contribution of point scatterers, the degree of polarization can increase to a significant extent [2]. Diffuse scattering is the main cause of depolarization and nonstationarity, especially in vegetated areas.

Results from the analysis and measurements of the long-term polarization behavior of ground clutter show similarities with target behavior [2]. When using linear polarization in the X band, these measurements show that the co-polarization return is greater than the cross-polarization one by about 6 dB on average, for both vertical and horizontal polarizations. This is typical for volume scattering from vegetation. When using circular polarization, the power of the cross-polarization return usually slightly prevails over that of the co-polarization return. Measurements carried out under this condition resulted in an average ratio, between cross-polarization and co-polarization power, ranging from 2 to 4 dB. The reader is referred to [55] and the bibliography therein for a detailed treatment of the subject.

13.4.4 Sea Clutter

When a radar from above illuminates an object on sea surface such as a ship, significant backscattering from the sea surface limits the target detection capability of the radar. For smooth sea, the backscattering appears to emanate from coherent sources located on the surface. In practice, one encounters rough seas, and multiple interactions of the electromagnetic wave with the ocean surface contribute incoherent random point scatterers to the return, similar to the diffuse scattering observed for land clutter. Measurements and theory indicate that sea clutter depends on radar frequency, polarization, grazing angle of the incident wave, wind speed, and sea state. For low grazing angles and at lower microwave frequencies (L-band or lower), it has been found that horizontal polarization offers considerably smaller radar cross section than vertical polarization. For rough seas, at lower grazing angles and higher frequencies, this difference is quite small, and even nonexistent. Measurements on sea clutter with other types of polarization, such as circular, have not been considered mostly because such measurements cannot be carried out under controlled conditions, and produce large variability in measured RCS. In spite of these difficulties, linear horizontal polarization is used to reduce sea clutter at lower grazing angles. This choice of polarization is advantageous for ship-mounted radars to target coastline and other surface vessels, as the target detectability is considerably enhanced, especially at the lower microwave frequencies in the L-band [56]. Ship detection

is becoming a major application of spaceborne SAR. HH is preferred for ship detection at high incidence angle (low grazing angles) and HV is preferred at low incidence (large grazing angles). Wetzel [57] discusses scattering models for sea clutter and summarizes the state-of-the-art in sea clutter characterization.

13.5 | MEASUREMENT OF THE SCATTERING MATRIX

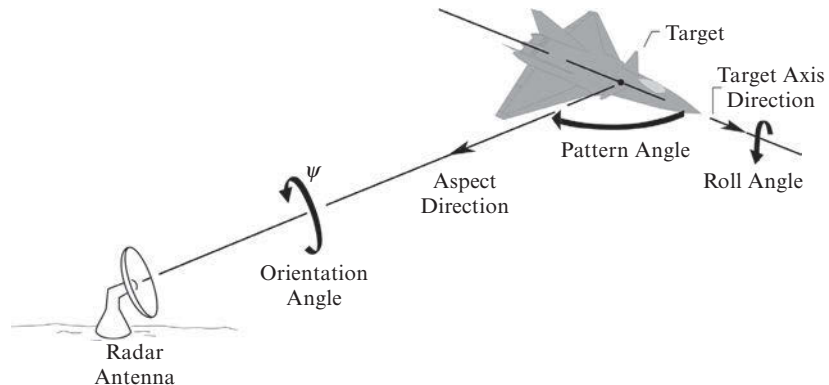
Polarimetric techniques enhance radar system performance by providing target signatures of higher fidelity and directly aid in minimizing the effects of clutter in the environment. The underlying principle of polarimetry requires the relative comparison of two orthogonal channels in amplitude and phase from the scattering of two successive polarization states transmitted by the radar. Building a database of target aspect angles for comparison must first be performed to adequately profile the target. In this section, measurement of the polarization scattering matrix will be discussed, considering a high-level block-diagram description of the necessary equipment and instrumentation setup. The system described features two orthogonally polarized receive channels alternately switching between two orthogonal polarization states in transmission [35]. This kind of acquisition allows the measurement of the entire scattering matrix with two sequentially transmitted pulses. There exist techniques for true simultaneous measurement of the complete target scattering matrix, without the need of a polarization switch [58]. These techniques require the simultaneous transmission of a pair of pulsed signals with orthogonal polarization states. Subject to ideal orthogonality conditions for the transmitted signals, all the target scattering matrix elements are simultaneously retrieved through a four-channel cross-correlation receiver.

The polarization response of a target is completely determined by the four-element scattering matrix in (13.31), rewritten to emphasize that the scattering dyadic is a function of target observation angle and radar frequency.

$$S(\theta_{\text{aspect}}, \theta_{\text{orient}}, f) = \begin{bmatrix} S_{HH}(\theta_{\text{aspect}}, \theta_{\text{orient}}, f) & S_{VH}(\theta_{\text{aspect}}, \theta_{\text{orient}}, f) \\ S_{HV}(\theta_{\text{aspect}}, \theta_{\text{orient}}, f) & S_{VV}(\theta_{\text{aspect}}, \theta_{\text{orient}}, f) \end{bmatrix} \quad (13.60)$$

As shown in Figure 13-13 [35], the aspect angle is defined as the angle that the radar line of sight makes with the target axis, and the orientation angle is the target rotation about the

FIGURE 13-13 ■ Target aspect direction and orientation angle. (From Huynen [35]. With permission.)



line of sight. One may define a target-centered Cartesian coordinate system in which the target axis is the z-axis, so that the roll angle defines the azimuth ϕ , and the aspect angle becomes θ . For monostatic RCS measurements, however, it is convenient to use the radar coordinate system as the reference [3]. The scattering matrix in (13.60) assumes a linear polarization basis; it can be transformed to circular polarization basis using a simple linear transformation (cf. (13.32)). In sequential measurement of the scattering matrix, linear vertical and horizontal polarization signals are transmitted in sequence, and both co-polarized and orthogonally polarized returns from the target are received. Reciprocity implies that $S_{HV} = S_{VH}$. Furthermore, for rotationally symmetric targets, if the line of sight or aspect direction is aligned with the target axis of symmetry, then the orientation angle becomes the roll angle, and it turns out that $S_{HV} = S_{VH} = 0$, and $S_{HH} = S_{VV}$.

As described in Section 13.3, there are two ways of characterizing scattering: (a) the complex voltage-type scattering matrix \mathbf{S} in which the absolute phase of the target return (or each matrix element) is needed; and (b) the real power-type Kennaugh matrix \mathbf{K} , which does not depend on the absolute target phase. Only the measurement of \mathbf{S} is addressed in this chapter. The reader is referred to [35] for measurement of the Kennaugh matrix. The absolute phase of a target depends on its local position, aspect, surface structure, and the radar frequency. Relative phases, on the other hand, are phase differences among the individual scattering coefficients where any one may act as phase reference for the others. The scattering matrix with absolute phase (SMA) is sensitive to target displacement along the line of sight, while the scattering matrix with relative phase (SMR) is not. From a measurements point of view, the difference between the two scattering matrix types is profound. Determination of the SMA requires the ability to measure absolute phase, while the SMR can be obtained by amplitude and relative phase measurements only. Relative phase measurements provide a means to achieve background subtraction of the environmental effects either experimentally (using a CW nulling technique) or computationally (using vector subtraction derived from auxiliary measurements) [35]. In power- or amplitude-only measurements, such as the Kennaugh matrix, the target returns for several transmit antenna polarization states are considered [35]. In conventional RCS measurements, just like in radar, the scattering matrix is measured by considering both amplitude and phase of a linearly polarized transmit channel, and only amplitude of the linearly polarized receive channel. Furthermore, by reconfiguring the transmit antenna polarization, all the important parameters of the target return can be measured [14]. This measurement, termed by Huynen as the linear restricted scattering matrix (LSM) [35], should not be confused with Kennaugh matrix, although the latter, in principle, can be derived from the former. Measurement of SMR is addressed next.

The direct method of measuring the scattering matrix effectively implements (13.60). The measurement of SMR requires two orthogonal transmitter polarizations being radiated independently in sequence, while amplitude and phase of the scattered return for each are observed simultaneously on dual-polarized orthogonal receiver channels. In general, any arbitrary pair of orthogonal illuminating polarizations can be used (e.g., right and left circular polarizations as well as vertical and horizontal). Instrumentation requirements include the ability to measure both amplitude and phase of the radar return, the capability of transmitting two orthogonally polarized signals in sequence, and that of receiving and recording the orthogonal components of the elliptically polarized returns simultaneously. For measurement of returns in low signal-to-noise ratio (SNR) environments or to minimize the effects of multipath and clutter, it is important to provide high isolation between the radar transmit and receive channels. This requires low-loss filters, circulators, and

multiplexers, among others, and considerably increases the instrumentation complexity. Other factors include good balance and isolation between the H and V receive channels. Calibration of the polarimetric measurements requires reference targets and a method to derive the polarimetric distortion matrices on transmit and receive.

Monostatic conditions for laboratory measurement are approximated by placing the transmit and receive antennas in proximity—a difficult performance trade-off considering that the mutual coupling increases with proximity, while absolute phase error, introduced by angular distortion arising from the quasi-bistatic positioning, decreases. In contrast, the direct measurement of the scattering matrix in the field can provide adequate separation between transmit and receive antennas for full gating of time domain signals, enabling a common aperture to be used for both channels. This eliminates the phase error by inherently setting the bistatic angle to zero, and backscattering from the environment is usually eliminated [35]. The received signal co-polarized with the transmit channel serves as the common phase reference in the measurement of SMA. This leads to relative phase estimate in the cross-polarized channel for polarimetry. Therefore, the acquisition of full target matrix requires returns of two orthogonal transmit polarization states to be measured on both horizontal and vertical receivers, with co-polarized transmitted signal serving as common phase reference.

A block diagram of a radar system for direct measurement of SMR is shown in Figure 13-14 [35]. Amplitude and phase data are obtained by using a coherent oscillator (COHO), a stable oscillator locked to the intermediate frequency (IF) in the receiver. On transmit, the output of the COHO is mixed with the frequency of a stable local oscillator (STALO) and is upconverted to a radio frequency (RF) that is equal to the sum of the COHO and STALO frequencies. This RF signal is amplified by a pulse-modulated power amplifier before transmission—it is thus coherent with the COHO reference signal. A dual-mode (orthomode) transducer is commonly used to provide a dual-polarized feed for exciting two orthogonal polarization states in the antenna. Dual linear or circular polarizations are thus generated for antenna illumination. Dual linear polarization is produced naturally by the dual-mode transducer, whereas circular polarization is effected by amplitude and phase control in the feed network using precision attenuators and phase shifters. The hybrids are essentially duplex filters that allow the same antenna element to be used for transmission and reception. For each linear polarization transmitted, co-polarized and cross-polarized signals are received and simultaneously processed in the appropriate channels.

The RF echo from the target is heterodyned with the STALO signal to produce the intermediate frequency. This signal is applied to a phase and an amplitude detector, as is the reference signal from the COHO. The output from the phase detector is proportional to the phase difference between the two input signals, whereas the output of the amplitude detector is proportional to the target echo signal referenced to the COHO signal. Calibration of the reference COHO signal in terms of the RCS of a calibration object allows us to characterize distortion matrices for transmit and receive channels, which are then applied to the measured target scattering matrix to compensate for polarimetric errors.

To obtain the scattering matrix of a given radar target, the amplitude and phase data for horizontal and vertical polarization combinations must be measured. This is accomplished by the use of two pulse-modulated power amplifiers, alternately pulsed on and off, feeding the vertically and horizontally polarized antennas. Scattering matrix data can be gathered on a two-pulse basis by exciting the vertically polarized antenna on one pulse, the horizontally polarized antenna on the next pulse, each time obtaining received vertically and horizontally polarized amplitude and phase data. When the target has been detected and

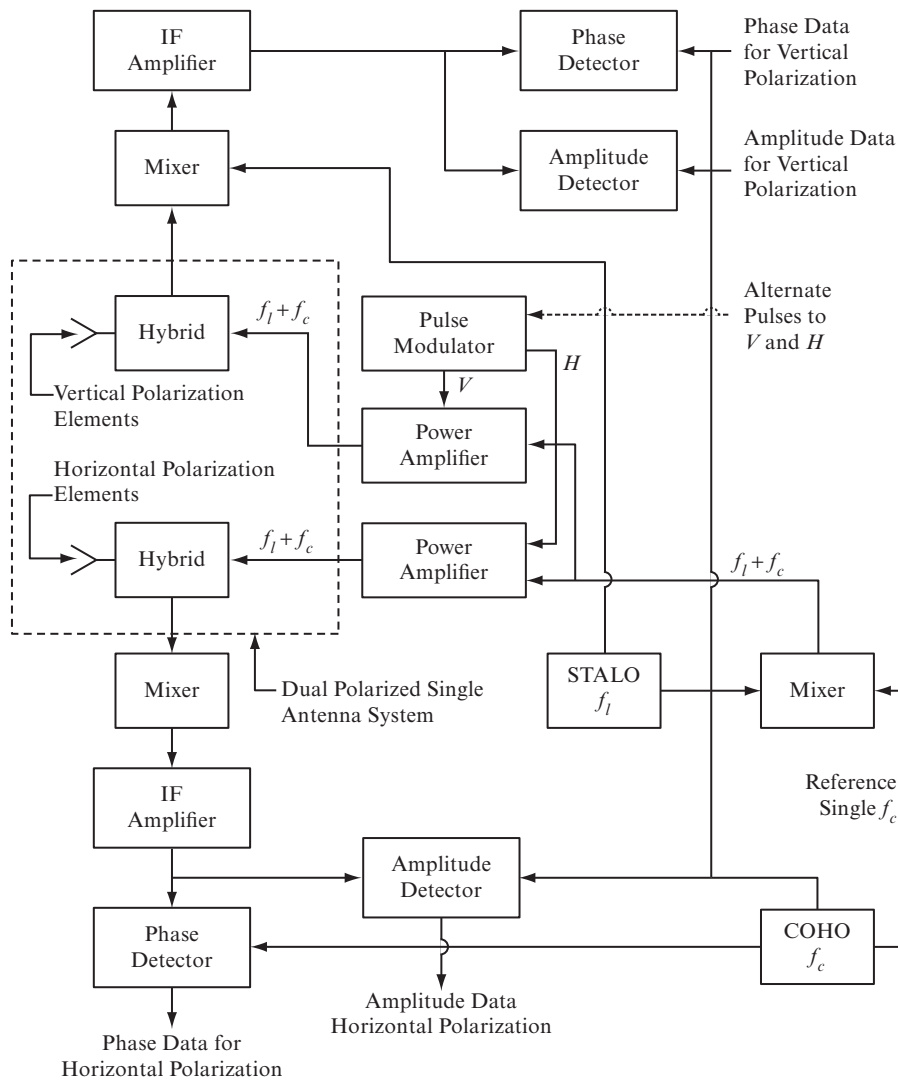


FIGURE 13-14 ■ A radar system for direct measurement of the scattering matrix. (From Huynen [35]. With permission.)

the pulse sequence established, only one additional pulse is required to provide scattering matrix data, since the previous pulse is used as a reference and contains half of the desired matrix information.

The white paper “New Network Analyzer Methodologies” [59] discusses how advances in network analyzer hardware and software provide increased functionality, speed and accuracy for scattering matrix measurements. The procedure for the measurement of a linear-basis full polarization matrix is described. This requires the source to transmit vertical and horizontal polarized fields and simultaneously measure received cross and co-polarized states. The paper describes how a programmable network analyzer can be used to eliminate external transmit polarization switch, and describes independent internal control of the analyzers, receivers and the source. The cross-polarized and co-polarized received signals can be simultaneously measured during each transmit polarization state because the analyzer processes these two states in the A and B receivers at the same time.

13.6 | SUMMARY

Wave polarization and polarimetry play vital roles in radar target identification comprising the three phases of target detection, discrimination, and recognition. This chapter presents an introduction to radar polarimetry emphasizing the polarization behavior as described by the scattering matrix. After a brief review of linear, circular, and elliptical polarization and the geometrical parameters of polarization ellipse, such as tilt and ellipticity angles and axial ratio, Stokes parameters and scattering matrices are introduced and explained using examples on canonical targets. The projection of polarization states on the Poincaré sphere is discussed. The transformation of the scattering matrix from one polarization state basis to another (e.g., linear to circular and vice versa) is derived. Unitary transformations are applied to the scattering matrix for fully polarized targets and eigen-polarization states corresponding to maximum and minimum backscattered power are derived. The location of null polarization states on the Poincaré sphere is discussed, and the Huynen polarization fork is introduced as a useful tool in the visualization of optimal polarimetric parameters. Stokes reflection or Mueller matrix is derived for partially polarized waves. The distinction between Kennaugh's and Huynen's formulations of the target scattering matrix (partially polarized case) is discussed. To demonstrate the application of radar polarimetry, the scattering matrix measurements of a cone with grooves (resembling a missile reentry vehicle) are processed to extract key features such as specular scattering and edge contributions, and the usefulness of polarimetric signatures to discriminate between coated and uncoated bodies is discussed. The polarimetric behavior of precipitation clutter, sea clutter, and ground clutter is also discussed. A radar instrumentation setup for scattering matrix measurement in block diagram form is described, and implications on the setup imposed by colocation of transmit and receive antennas for monostatic measurements and the need to measure amplitude and phase of each orthogonal channel relative to a coherent source are discussed.

13.7 | FURTHER READING

This chapter provided an introduction to radar polarimetry based on the original theory developed by Kennaugh [13] and Huynen [14]. Significant modern developments in radar polarimetry and its application to imaging have been described in the papers, books and monographs listed below.

- Boerner and Lee [60] provide a succinct review of the existing pertinent monographs, books, and study guides on radar polarimetry, and identify particular topics that are being covered in new books in order to reflect on complexity of state-of-the-art polarimetric systems as well as modern evolution of analytical techniques for polarimetry beyond the seminal methods introduced by Kennaugh.
- Boerner, Wolfgang-Martin, *Introduction to Synthetic Aperture Radar (SAR) Polarimetry*, Wexford Press, April 2007. Available from: <http://www.rto.nato.int/abstracts.asp>
- Yamaguchi, Yoshio, *Radar Polarimetry from Basics to Applications: Radar Remote Sensing using Polarimetric Information* (in Japanese), IEICE Press, Dec. 2007.
- Masonnett, Didier and Souyris, Jean-Claude, *Imaging with Synthetic Aperture Radar*, CRC Press, Taylor and Francis, 2008.

- Lee, Jong-Sen and Pottier, Eric, *Polarimetric Radar Imaging – From Basics to Applications*, CRC Press, Taylor and Francis, 2009.
- Cloude, Shane Robert, *Polarisation: Applications in Remote Sensing*, Oxford University Press, 2009.
- VanZyl, Jakob-Johannes and Kim, Yun-Jin, *Polarimetric SAR Imaging*, John Wiley and Sons, New York, 2010.

Acknowledgment: The author is grateful to Dr. Jean Piou, at MIT Lincoln Laboratory, who introduced him to polarimetric radar imaging and target feature extraction using spectral estimation methods. Constructive comments from a thorough revision by an anonymous reviewer served to improve the clarity of presentation.

13.8 | REFERENCES

- [1] Poelman, A.J. and Guy, J.R.F., “Polarization Information Utilization in Primary Radar,” in *Direct and Inverse Methods in Radar Polarimetry*, Ed. W.-M. Boerner et al., Parts 1 and 2, Kluwer Academic Publishers, Netherlands, 1992.
- [2] Giuli, D., “Polarization Diversity in Radar,” in *Proceedings of the IEEE*, vol. 79, no. 2, pp. 245–269, 1986.
- [3] Mott, H., *Antennas for Radar and Communications – A Polarimetric Approach*, John Wiley, New York, 1992, Chapter 6.
- [4] Ezquerro, N.F., “Target Recognition Considerations,” in *Principles of Modern Radar*, Ed. J.L. Eaves and E.K. Reedy, Van Nostrand Reinhold, New York, 1987.
- [5] Rihaczek, A.W. and Hershkowitz, S.J., *Theory and Practice of Radar Target Identification*, Artech House, Norwood, MA, 2000.
- [6] Skolnik, M.I., Ed., *Radar Handbook*, 2d ed., McGraw-Hill, New York, 1990.
- [7] Ulaby, F.T., Moore, R.K., and Fung, A.K., *Microwave Remote Sensing*, vols. 1–3, Addison-Wesley, Reading, MA, 1981.
- [8] Ulaby, F.T. and Elachi, C., Eds., *Radar Polarimetry for Geoscience Applications*, Artech House, Norwood, MA, 1990.
- [9] Bringi, V.N. and Chandrasekar, V., *Polarimetric Doppler Weather Radar—Principles and Applications*, Cambridge University Press, Cambridge, UK, 2001.
- [10] Elachi, C., *Space-Borne Radar Remote Sensing: Applications and Techniques*, IEEE Press, New York, 1988.
- [11] Lee, J.-S. and Pottier, E., *Polarimetric Radar Imaging—From Basics to Applications*, CRC Press, Boca Raton, FL, 2009.
- [12] Kennaugh, E.M., “Polarization Properties of Radar Reflections,” M.Sc. thesis, Department of Electrical Engineering, The Ohio State University, Columbus, Rep. 389-12, March 1952.
- [13] Kennaugh, E.M., “Research Studies on the Polarization Properties of Radar Targets,” Collected Volumes of the Final Reports for USAF Contract No. 28(099)-90, Report No. 389-1 to 15, and 389-17 to 24, vols. 1 and 2, July 1984. (Available from Librarian, Ohio State University, ElectroScience Laboratory, 1420 Kinnear Road, Columbus, OH 43212.)

- [14] Huynen, J.R., "Phenomenological Theory of Radar Targets," Ph.D. dissertation, University of Technology, Delft, Netherlands, December 1970. Available online from <http://repository.tudelft.nl/view/ir/uuid%3Ae4a140a0-c175-45a7-ad41-29b28361b426/>
- [15] Graves, G.A., "Radar Polarization Power Scattering Matrix," *Proceedings of the IEEE*, vol. 44, pp. 248–252, 1956.
- [16] Maffett, A.L., "Scattering Matrices," *Methods of Radar Cross Section Analyses*, Ed. J.W. Crispin and K.M. Siegel, Academic Press, New York, 1968.
- [17] van Zyl, J.J., "On the Importance of Polarization in Radar Scattering Problems," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, January 1986.
- [18] Liu, C.-L., Zhang, X., Yamaguchi, Y., Eom, H.J., and Boerner, W.-M., "Comparison of Optimization Procedures for 2×2 Sinclair, 2×2 Graves, 3×3 Covariance, and 4×4 Mueller (Symmetric) Scattering Matrices in Coherent Radar Polarimetry and Its Application to Target vs. Background Discrimination in Microwave Remote Sensing and Imaging," in *Proceedings of SPIE*, vol. 1748, pp. 144–173, February 1993.
- [19] Kostinski, A.B. and Boerner, W.-M., "On Foundations of Radar Polarimetry," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 12, pp. 1395–1404, December 1986.
- [20] Born, M. and Wolf, E., *Principles of Optics*, 6th ed., Pergamon Press, Oxford, 1980.
- [21] Henderson, F.M. and Lewis, A.J., Eds., *Principles and Applications of Imaging Radar, Manual of Remote Sensing*, 3rd ed., vol. 2, John Wiley, New York, July 1998.
- [22] Hollis, J.S., Hickman, T.G., and Lyon, T., "Polarization Theory," in *Microwave Antenna Measurements Handbook*, Scientific Atlanta, Atlanta, GA, 1970.
- [23] Holm, W.A., "Polarimetric Fundamentals and Techniques," in *Principles of Modern Radar*, Ed. J.L. Eaves and E.K. Reedy, Van Nostrand Reinhold, New York, 1987.
- [24] Burrows, M.L., "Two-Dimensional ESPRIT with Tracking for Radar Imaging and Feature Extraction," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 2, pp. 524–531, February 2004.
- [25] Mayhan, J., Burrows, M.L., Cuomo, K.M., and Piou, J.E., "High Resolution 3D 'Snapshot' ISAR Imaging and Feature Extraction," *IEEE Transactions on Aerospace Electronic Systems*, vol. 20, pp. 630–642, April 2001.
- [26] Naishadham, K. and Piou, J.E., "Spectral Characterization of Extended Returns in Radar Target Signatures," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 6, pp. 1742–1751, June 2008.
- [27] Sinclair, G., "The Transmission and Reception of Elliptically Polarized Waves," *Proceedings of the IRE*, vol. 38, pp. 148–151, February 1950.
- [28] Davidovitz, M. and Boerner, W.-M., "Extension of Kennaugh's Optimal Polarization Concept to the Asymmetric Matrix Case," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 569–574, April 1986.
- [29] Agrawal, A.P. and Boerner, W.-M., "Redevelopment of Kennaugh's Target Characteristic Polarization State Theory Using the Polarization Transformation Ratio Formulism for the Coherent Case," *IEEE Transactions on Antennas and Propagation*, vol. AP-27, no. 1, pp. 2–14, January 1989.
- [30] Boerner, W.-M., Yan, W.-L., Xi, A.-Q., and Yamaguchi, Y., "On the Basic Principles of Radar Polarimetry: The Target Characteristic Polarization State Theory of Kennaugh, Huynen's Polarization Fork Concept, and Its Extension to the Partially Polarized Case," *Proceedings of IEEE*, vol. 79, no. 10, pp. 1538–1550, October 1991.
- [31] Balanis, C.A., *Antenna Theory: Analysis and Design*, 3rd ed., John Wiley, New York, 2005.

- [32] Kraus, J.D., *Electromagnetics*, 3rd ed., McGraw Hill, New York, 1984.
- [33] Jones, R.C., “A New Calculus for the Treatment of Optical Systems, I. Description and Discussion of the Calculus,” *Journal of the Optical Society of America*, vol. 31, no. 7, pp. 488–493, 1941.
- [34] Stokes, G., “On the Composition Resolution of Streams of Polarized Light from Different Sources,” *Transactions on Cambridge Philosophical Society*, vol. 9, no. 3, pp. 399–416, 1852.
- [35] Huynen, J.R., “Measurement of the Target Scattering Matrix,” *Proceedings of the IEEE*, vol. 53, pp. 936–946, August 1965.
- [36] Yamaguchi, Y., Boerner, W.-M., Eom, H.J., Sengoku, M., and Abe, T., “Characteristic Polarization States in the Cross-Polarized Radar Channel,” in *Proceedings of the SPIE*, vol. 1748, pp. 174–183, February 1993.
- [37] Boerner, W.-M. and Xi, A.-Q., “The Characteristic Radar Target Polarization State Theory for the Coherent Monostatic and Reciprocal Case Using the Generalized Polarization Transformation Ratio Formulation,” *Archiv der Elektrischen Übertragung (AEU)*, vol. 44, no. 4, pp. 273–281, July 1990.
- [38] Mott, H., *Remote Sensing with Polarimetric Radar*, IEEE Press and John Wiley, Hoboken, NJ, 2007.
- [39] Cuomo, K.M., Piou, J.E., and Mayhan, J.T., “Ultra-Wideband Coherent Processing,” *IEEE Transactions on Antennas and Propagation*, vol. 47, no. 6, pp. 1094–1107, June 1999.
- [40] Claude, S.R. and Pottier, E., “A Review of Target Decomposition Theorems in Radar Polarimetry,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 2, pp. 498–518, March 1996.
- [41] Huynen, J.R., “Physical Reality of Radar Targets,” in *Proceedings of the SPIE*, vol. 1748, pp. 86–96, February 1993.
- [42] Barnes, R.M., “Roll-Invariant Decompositions for the Polarization Covariance Matrix,” *Polarimetry Technology Workshop*, Redstone Arsenal, AL, 1988.
- [43] Holm, W.A. and Barnes, R.M., “On Radar Polarization Mixed State Decomposition Theorems,” in *Proceedings of the USA National Radar Conference*, April 1988.
- [44] Knott, E.F., Shaeffer, J.F., and Tuley, M.T., *Radar Cross Section*, 2d ed., SciTech, Raleigh, NC, 2004.
- [45] Mensa, D.L., *High Resolution Radar Cross Section Imaging*, Artech House, Boston, MA, 1991.
- [46] Proakis, J.G., Rader, C.M., Ling, F., and Nikias, C.L., *Advanced Digital Signal Processing*, Macmillan, New York, 1992.
- [47] Paulraj, A., Roy, R., and Kailath, T., “Estimation of Signal Parameters via Rotational Invariance Techniques—ESPRIT,” in *Proceedings of the 19th Asilomar Conference on Circuits, Systems and Computers*, Ed. D.E. Kirk, Pacific Grove, CA, pp. 83–89, November 1985.
- [48] Kung, S.Y., Arun, K.S., and Rao, D.V.B., “State-Space and Singular Value Decomposition-Based Approximation Methods for the Harmonic Retrieval Problem,” *Journal of the Optical Society of America*, vol. 73, pp. 1799–1811, December 1983.
- [49] Choi, J., Wang, N., Peters, Jr., L., and Levy, P., “Near Axial Backscattering from Finite Cones,” *IEEE Transactions on Antennas and Propagation*, vol. 38, no. 8, pp. 1264–1272, August 1990.
- [50] Naishadham, K. and Felsen, L.B., “Dispersion of Waves Guided along a Cylindrically Stratified Medium,” *IEEE Transactions on Antennas and Propagation*, vol. 41, no. 3, 304–313, March 1993.

- [51] Naishadham, K. and Piou, J.E., “Analytical Characterization and Validation of Creeping Waves on Dielectric Coated Cylinders,” *Radio Science*, vol. 45, RS5014, October 2010. doi:10.1029/2009RS004241.
- [52] Bean, B.R., Dutton, E.J., and Warner, B.D., “Weather Effects in Radar,” in *Radar Handbook*, Ed. M.I. Skolnik, McGraw Hill, New York, 1970, Chapter 24.
- [53] Baltan, L.J., *Radar Observations of the Atmosphere*, University of Chicago Press, 1973.
- [54] Serafin, R.J., “Meteorological Radar,” in *Radar Handbook*, ed. M.I. Skolnik, 2nd ed., McGraw Hill, New York, 1990, Chapter 23.
- [55] Moore, R.K., “Ground Echo,” in *Radar Handbook*, Ed. M.I. Skolnik, 2nd ed., McGraw Hill, New York, 1990, Chapter 12.
- [56] Croney, J., “Civil Marine Radar,” in *Radar Handbook*, Ed. M.I. Skolnik, McGraw Hill, New York, 1970.
- [57] Wetzel, L.B., “Sea Clutter,” in *Radar Handbook*, Ed. M.I. Skolnik, 2nd ed., McGraw Hill, New York, 1990, Chapter 13.
- [58] Giuli, D., Fossi, M., and Facheris, L., “Radar Target Scattering Matrix Measurement through Orthogonal Signals,” in *IEEE Proceedings*, Pt. F, vol. 140, no. 4, pp. 233–242, August 1993.
- [59] “New Network Analyzer Methodologies in Antenna/RCS Measurements,” White Paper, Agilent Technologies, Product No. 5989-1937EN, 2004.
- [60] Boerner, W.-M. and Lee, J.-S., “Review of Existing Monographs and Books on Radar Polarimetry and Polarimetric SAR with the Aim of Justifying the Need of Updates,” in *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, Barcelona, pp. 180–183, 2007.

13.9 | PROBLEMS

1. Write the expression for the electric field intensity of a linearly polarized wave traveling along the z -direction with $E_{0x} = 2$ V/m and $E_{0y} = 1$ V/m. Assume a phase shift $\delta = 0$ between the field components. Find the tilt angle ψ .
2. (a) Write an expression for the electric field of a right-handed circularly polarized wave with a frequency of 300 MHz and an amplitude of 2 V/m, traveling along the z -direction.
 (b) Assume that this wave is incident normally on a perfectly conducting wall, of infinite extent, located at $z = 0$. Determine the electric field intensity of the reflected wave. Determine the polarization type and rotation sense of the reflected wave.
 (c) Repeat (b) with the perfectly conducting wall replaced by a dielectric half-space with $\mu_r = 1$, $\varepsilon_r = 4$. Write expressions for the reflected and transmitted waves and determine their polarization as well as rotation sense.
3. An elliptically polarized wave is traveling along the z -direction with $E_{0x} = 2$ V/m, $E_{0y} = 3$ V/m, and a phase shift $\delta = -30^\circ$ between the field components. (a) Write the expression for the electric field intensity and determine the rotation sense of the polarization ellipse. (b) Determine the axial ratio, ellipticity angle, and tilt angle. (c) Calculate the polarization ratio P . (d) Determine the rectangular components of the field in terms of the tilt and ellipticity angles using (13.20) and show that it agrees

with the result in (a). Choose the reference angle α such that the x-component of the field is real.

4. An elliptically polarized wave is traveling along the z -direction with the phasor electric field given by $\mathbf{E} = (\hat{x}2 + \hat{y}1e^{j\pi/4})e^{-jkz}$. Find the tilt angle, ellipticity angle, and sense of rotation of the polarization ellipse. Find a complex unit vector to represent its polarization state in a linear ($x-y$) basis. It is required to represent this wave as a sum of two orthogonally polarized elliptical states, $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 = (\hat{e}_1 E_1 + \hat{e}_2 E_2)e^{-jkz}$. Assume that the field \mathbf{E}_1 has tilt angle $\psi = 30^\circ$ and ellipticity angle $\tau = -22.5^\circ$. Using the linear-to-elliptical polarization basis transformation in (13.35), calculate E_1, E_2 , and determine the complex vectors for the orthogonal elliptical polarizations states \hat{e}_1 and \hat{e}_2 .
5. Write the elliptically polarized wave of Problem 4 as the sum of two orthogonal circularly polarized waves represented by the basis vector $[\hat{e}_r \hat{e}_{r\perp}]$ (see (13.30)).
6. An elliptically polarized wave is traveling along the z -direction with the phasor electric field given by $\mathbf{E} = (\hat{x}5 + \hat{y}8e^{j\pi/8})e^{-jkz}$. Find the tilt angle, ellipticity angle, and sense of rotation of the polarization ellipse. Find the Stokes parameters and determine the location of the polarization state on the Poincaré sphere. Also, determine the antipodal polarization state and write its Stokes vector.
7. If the Stokes parameters of a wave are given by $g_0 = 3, g_1 = -2, g_2 = 1, g_3 = 2$, find E_x and E_y , the tilt angle and the ellipticity angle. Show the location of its polarization state on the Poincaré sphere.
8. Consider a wave with field components given by

$$E_x = 2 \cos\left(\omega t + \frac{\pi}{8}\right)$$

$$E_y = 3 \cos\left(\omega t + \frac{3\pi}{2}\right).$$

Find the Stokes parameters and their Poincaré representation (13.21) in terms of tilt and ellipticity angles. Show the polarization state location on the Poincaré sphere.

9. A partially polarized wave for which $d = 0.5$, $AR = 2$, $\psi = 0^\circ$, and $S = 1$ (W/m²), where S is the average power density of the wave, is incident on an antenna for which $AR = 4$, $\psi = 22.5^\circ$, and the effective aperture $A_e = 1$ m². Find the received power.
10. (a) A wave for which $d = 0.4$, $AR = 3$, $\psi = 45^\circ$, and $S = 1$ (W/m²) is separately received by six antennas, all of unit effective aperture, with polarization as follows: (i) linear horizontal, (ii) linear vertical, (iii) linear slant (45°), (iv) linear slant (135°), (v) left circular, and (vi) right circular. In each case, find the received power.
 (b) Is there a wave to which all the six antennas listed in part (a) respond equally? If so, what are the wave parameters (tilt angle and axial ratio)?
11. Show that a very thin square plate or disk of area A_p has the scattering matrix in linear polarization basis given by

$$\mathbf{S} = C \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, C = \frac{2\sqrt{\pi}A_p}{\lambda}$$

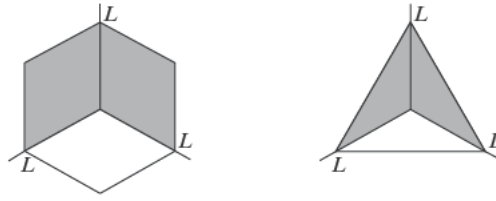


FIGURE 13-15 ■ Trihedral corner reflectors: (a) Square. (b) Triangular. (From Mott [3]. With permission.)

where the amplitude C is obtained at specular aspect using the physical optics (PO) approximation [44], and λ is the wavelength. The radar cross section for linear polarization is thus given by $\sigma = |C|^2$. It is convenient to ignore the amplitude unless one is calculating absolute power. Show that the five Huynen fork parameters are given by

$$m = C, \psi = \text{arbitrary}, v = 0^\circ, \tau = 0^\circ, \gamma = 45^\circ.$$

Note that for the flat disk the rotation of the target about the radar line of sight aspect does not alter the received power, and therefore, the orientation angle ψ is arbitrary.

12. Show that a sphere has the scattering matrix given by

$$S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and the five Huynen fork parameters, by $m = C$, $\psi = \text{arbitrary}$, $v = 0^\circ$, $\tau = 0^\circ$, $\gamma = 45^\circ$. The amplitude m follows from the PO approximation [44].

13. Figure 13-15 depicts square and triangular trihedral corners with edge length L along each coordinate direction [3]. Show that the trihedral corner has the same scattering matrix and Huynen fork parameters as those for the sphere, except for the amplitudes m , which are given by [3]

$$m = \begin{cases} 2\sqrt{3\pi}L^2/\lambda, & \text{Square corner} \\ 2\sqrt{\pi}L^2/\lambda, & \text{Triangular corner} \end{cases}$$

14. Show that the scattering matrix for a helix is given by

$$S = \frac{1}{2} \begin{bmatrix} 1 & \pm j \\ \pm j & -1 \end{bmatrix}$$

where the $+$ sign in the off-diagonal terms pertains to a left-handed helix, and the $-$ sign to a right-handed helix. Since the helix is invariant to rotations about the line of sight, the orientation angle ψ is arbitrary. By diagonalization of the scattering matrix, show that within constant amplitude the Huynen fork parameters are given by $\psi = \text{arbitrary}$, $v = \text{arbitrary}$, $\tau = \pm 45^\circ$, $\gamma = 0^\circ$. It is evident that the wave scattered by the helix is circularly polarized, with the right-handed helix producing RCP ($\tau = -45^\circ$), and the left handed helix, LCP ($\tau = +45^\circ$).

15. A left-circularly polarized wave from a radar is scattered from a target, with its scattering matrix given by

$$S = \begin{bmatrix} 2 & j1 \\ j1 & 4 \end{bmatrix}$$

Diagonalize S and determine the two eigenvalues λ_1 and λ_2 in (13.36). Find the transmit and receive polarization ratios of a monostatic radar antenna to (a) maximize and (b) minimize the copolarized received power. Find the polarization states (complex unit vectors) corresponding to the polarization ratios, and show that the state vectors are orthonormal (see Chapter 6 in [3]).

16. Assume that the scattering matrix of a target is given by

$$S = \begin{bmatrix} 2j & 0.5 \\ 0.5 & j \end{bmatrix}$$

Diagonalize S and determine the two eigenvalues λ_1 and λ_2 in (13.36). Using the three-step procedure proposed in [19], determine the optimal transmit and receive polarization states for maximum backscattered power.

17. A dihedral corner has the monostatic scattering matrix given by

$$S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Find the Kennaugh matrix. Show the optimal polarization states by locating the vertices of the polarization fork on the Poincaré sphere.

18. Determine the Kennaugh matrix for the spherical target, whose scattering matrix is derived in Problem 12.
19. Determine the Kennaugh matrices for the right- and left-handed helices, whose scattering matrices are derived in Problem 14.

Automatic Target Recognition

Kristin F. Bing, Lisa M. Ehrman, Teresa M. Selee

Chapter Outline

14.1	Introduction	631
14.2	Unified Framework for ATR	633
14.3	Metrics and Performance Prediction	634
14.4	Synthetic Aperture Radar	638
14.5	Inverse Synthetic Aperture Radar	652
14.6	Passive Radar ATR	656
14.7	High-Resolution Range Profiles	658
14.8	Summary	661
14.9	Further Reading	661
14.10	References	662
14.11	Problems	668

14.1 INTRODUCTION

The three main functions of radar are detection, tracking, and target recognition. Detection is discussed at length in *Principles of Modern Radar: Basic Principles*, and tracking is addressed in Chapter 15 of this text. This chapter focuses on target recognition.

Automatic target recognition (ATR) systems are deployed in a broad spectrum of military, civilian, and commercial applications, where they assist with everything from identifying aircraft [100–105] to spotting tornados [106] to countering bioterrorism [107–109]. Regardless of the application, the goal is the same. ATR systems attempt to automatically categorize targets into one of a set of classes. The sensors exploited in ATR systems are as varied as the applications. These include imaging sensors such as synthetic aperture radar (SAR), inverse synthetic aperture radar (ISAR), and electro-optical/infrared (EO/IR) focal plane arrays, as well as non-imaging sensors such as high range resolution (HRR) radar, bistatic passive radar, and Raman spectrometers [107–109].

Before progressing any further, it is worth noting that the terminology varies substantially from one radar application to another. For example, literature on SAR and ISAR typically use discrimination to refer to the process of segmenting an image into the broad classes of (natural) background and man-made portions. The subsequent process of assigning a target to a specific category of interest is then referred to as classification. However,

other parts of ATR literature sometimes refer to classification as the assignment of a target to a broad class of targets (e.g., military aircraft vs. commercial aircraft), while discrimination, identification, and target recognition are used interchangeably to refer to the assignment of a target to a more specific target class (e.g., F-4 vs. F-15). For consistency's sake, we have chosen to adopt the terminology associated with SAR literature.

14.1.1 Organization

A single chapter cannot possibly address the entire aforementioned body of ATR research. Hence, in keeping with the theme of this overall textbook, this chapter limits the scope to radar-based ATR schemes using sources found in open literature. Section 14.2 provides a unified framework for understanding basic ATR functionality; metrics for scoring ATR algorithms are presented in Section 14.3. The unified framework for ATR is then applied in Section 14.4 to SAR, in Section 14.5 to ISAR, in Section 14.6 to passive radar, and in Section 14.7 to high-resolution range profiles. Section 14.8 concludes with a summary.

14.1.2 Key Points

- Regardless of the application, most ATR algorithms can be decomposed into four main steps, which we have named the unified framework for ATR. The four steps include (1) identify the target set, (2) select the feature set, (3) observe the feature set, and (4) test the feature set.
- Although academic papers often focus on the last two steps, the algorithm will not be robust if the feature set does not adequately segregate the desired target set.

14.1.3 Acronyms

ATR	automatic target recognition
CFAR	constant false alarm rate
DOA	direction-of-arrival
DCCF	distance classifier correlation filter
EO/IR	electro-optical/infrared
FOA	focus of attention
HDVI	high-definition vector imaging
HMM	hidden Markov model
HRR	high-resolution radar
HRRP	high-resolution range profile
ICA	independent component analysis
IPR	image point responses
ISAR	inverse synthetic aperture radar
KF	Kalman filter
LDA	linear discriminant analysis
MACH	maximum average correlation height
MRA	multi-resolution analysis
MSE	mean-squared error
MSTAR	moving and stationary target acquisition and recognition
MTI	moving target indication
PCA	principal component analysis

PD	probability of detection
PEMS	predict, extract, match, and search
ROC	receiver operating characteristic
SAR	synthetic aperture radar
SVD	singular value decomposition
SVM	support vector machine
UHF	ultra high frequencies

14.2 | UNIFIED FRAMEWORK FOR ATR

Even though the applications of ATR are diverse, the basic framework for ATR is remarkably similar and can be decomposed into four main steps, referred to here as the unified framework for ATR. The following sections describe these steps, as well as common challenges.

14.2.1 Step 1: Identify the Target Set

The first step in development of an ATR algorithm is to identify the target set, which includes the targets of interest, as well as targets that are not of interest but are likely to be observed. Implicit in this decision is the desired level of precision. For example, the goal of some ATR systems is to distinguish between broad classes of targets, while the goal of others is to distinguish specific models within a class. In the former case, the target set may include broad classes such as aircraft, satellites, missiles, and unknown targets, whereas in the latter case it might include more precise classes of aircraft models such as F-15, F-16, T-38, and unknown targets. Subsequent decisions are heavily influenced by this first step.

14.2.2 Step 2: Select the Feature Set

Having identified the target set, the next step is to select a feature set that maximizes “the similarity of objects in the same class while maximizing the dissimilarity of objects in different classes” [118, pp. 367]. Sensitivity is a key component in this definition. Some features (e.g., radar cross section, when the wavelength is much smaller than the target) are extremely sensitive to small changes in target parameters like aspect or pose; hence, they do not maximize the similarity of objects in the same class, or even of the same object over time. The goal is to find a set of features sufficiently sensitive to reliably separate targets into classes but not so sensitive that targets in the same class are appropriated into disparate classes.

14.2.3 Step 3: Observe the Feature Set

Once the feature set has been selected, the next challenge is to observe it as accurately as possible. Depending on the application and the choice of features, this may imply that specialized signal processing or conditioning (e.g., background removal) is necessary to extract observations of the feature set from the collected data. Accurately observing the feature set may also require the sensor to operate in a non-standard mode (i.e., to use different waveforms than are normally applied), which could drive fielding decisions.

14.2.4 Step 4: Test the Feature Set

The final step assesses the observed feature set and identifies the target as belonging to a particular class of targets, determined in Step 1. A wide variety of algorithms are described in the open literature for testing the feature set. These will be presented in the context of typical applications later in the chapter.

14.2.5 Challenges

Certain challenges are implicit in each of the four steps in the Unified Framework for ATR. One of the primary challenges in identifying the target set is determining the range of variations that are likely. This obviously requires some prior knowledge of the target set, but even when the risk of unknown target classes is small, the possible variations for known targets can still be difficult to predict. Consider a tank. Its weapons may present in varying levels of articulation, it may be coated in varying levels of dirt after time in the field, or it may be partially obscured (perhaps intentionally) in foliage. Aircraft are similarly subject to wide variations; a variety of payloads and antennas may be possible for a given aircraft, each of which could substantially alter features such as radar cross section. Since each variation of a target could present substantially different features, the range of likely tactical conditions must be addressed if the ATR scheme is to be successful. As a result, identifying the target set can include identifying the range of probable conditions under which each target in the set may be observed.

As the range of probable presenting conditions increases, feature set selection becomes increasingly difficult. In fact, selecting an adequate feature set is often the most challenging and seminal part of the process. If a feature set that sufficiently segregates the target set (including all likely target variations) cannot be identified, little can be done to compensate in subsequent steps. Hence, feature set selection tends to drive critical characteristics of the ATR system, such as the sensor choice, the operational timeline, the number of targets that can be simultaneously identified, and the level of sophistication required when testing the feature. Predicting performance based on the available features can help in execution of this step; however, efficient and reliable performance prediction techniques for N -dimensional feature sets are not straightforward.

14.3 | METRICS AND PERFORMANCE PREDICTION

Subsequent sections in this chapter discuss a variety of radar-based ATR applications and the associated challenges. Before forging ahead, it is helpful to introduce application-independent mechanisms for scoring ATR algorithms.

14.3.1 Common Metrics

Two common metrics for assessing the performance of ATR schemes are receiver operating characteristic (ROC) curves and confusion matrices, both of which reflect the probability of detection (i.e., correct classification) and probability of false alarm (i.e., mis-classification).

ROC curves plot the probability of detection (PD) as a function of the probability of false alarm (PFA) for targets of interest. They can also serve as a useful design tool, inasmuch as they can illuminate the degree of improvement garnered from each step in the ATR scheme. PFA tends to decrease in Step 3 (observe the feature set), as increasingly

TABLE 14-1 ■ Notional Confusion Matrix

	ATR Target 1	ATR Target 2	...	ATR Target N
True Target 1	$c_{(1,1)}$	$c_{(1,2)}$		$c_{(1,N)}$
True Target 2	$c_{(2,1)}$	$c_{(2,2)}$		$c_{(2,N)}$
...				
True Target N	$c_{(N,1)}$	$c_{(N,2)}$		$c_{(N,N)}$

sophisticated techniques for feature extraction and observation are implemented, while PD gains are generally realized in Step 4 (test the feature set), as increasingly advanced tests are applied to pre-processed feature sets [1].

Confusion matrices essentially display snapshots of the same information in tabular form, as shown in Table 14-1. The ATR answers appear across the top row and the true target list appears in the leftmost column. Hence, the (i, j) element of the confusion matrix corresponds to the probability that target i is labeled target j by the ATR algorithm. The more the confusion matrix resembles the identity matrix, the better the results [2].

If Monte Carlo trials of simulated or real data are available, then the confusion matrices can be found by computing the fraction of Monte Carlo trials in which target i is labeled as target j . Furthermore, as long as the feature sets are statistically well-characterized, then performance prediction tools can predict the confusion matrices, alleviating the need for extensive Monte Carlo analysis.

This is particularly helpful in the design process. If the off-diagonal terms in the predicted confusion matrix are sufficiently small for the application, this justifies the ATR scheme design. If they are too large, then the ATR scheme design should be revisited. Maybe the observations are not accurately measured, preventing them from matching the predictions. Perhaps the chosen feature set simply fails to separate the target set reliably. If other features are available, then their inclusion via multi-dimensional analysis may solve the problem; otherwise, the granularity of the ATR answer may need to be adjusted.

14.3.2 Performance Prediction

Two general performance prediction metrics are available, both of which flow from hypothesis-testing. If the application is well-suited for the Neyman-Pearson framework, then the Type II error bounds (β) may be approximated from the relative entropy (i.e., Kullback-Leibler distance) [110]. Similarly, if the application is well-suited for the Bayesian context, then the probability of error bounds (P_E) may be approximated from the Chernoff information [110].

14.3.2.1 Performance Prediction in the Neyman-Pearson Framework

The Neyman-Pearson framework applies most naturally to applications in which one type of error has more dire consequences than the other. Consider an ATR scheme that supports fire-control decisions, where the null hypothesis is that a given target is friendly and the alternate hypothesis is that it is an enemy meriting weapon fire. Only a small probability of friendly fire (i.e., firing upon a friendly target) may be acceptable. For a given null hypothesis and alternate hypothesis, the acceptance threshold is then set to keep the probability of a Type I error (α) - incorrectly classifying a friendly target as an enemy - within the desired bounds. Since the probabilities of Type I and Type II errors cannot be simultaneously minimized, this implies that the resulting Type II error will be tolerated.

Under the Neyman-Pearson framework, Stein's lemma [110] states that bounds on the probability of a Type II error (which is equal to the off-diagonal entries in the confusion matrix) are approximated via the relative entropy (i.e., Kullback-Leibler distance, D). Hence, the Type II error between targets i and j (i.e., the probability that target j is incorrectly identified as target i) is given by

$$\beta_{i,j} = \exp(-D_{i,j}) \quad (14.1)$$

where

$$D_{i,j} = \int_0^\infty p_i(x) \ln \left(\frac{p_i(x)}{p_j(x)} \right) dx \quad (14.2)$$

and $p_i(x)$ is the probability density function of target i . If both target features are Gaussian, for example, with $p_i(x) \sim N(\mu_i, \sigma_i^2)$, then it is straightforward to show that

$$D_{i,j} = \ln \left(\frac{\sigma_i^2}{\sigma_j^2} \right) + \frac{1}{2\sigma_j^2} [\sigma_i^2 + (\mu_i - \mu_j)^2] - \frac{1}{2} \quad (14.3)$$

and

$$\beta_{i,j} \approx \left(\frac{\sigma_i^2}{\sigma_j^2} \right) \exp \left\{ -\frac{1}{2} \left[\frac{\sigma_i^2}{\sigma_j^2} + \frac{(\mu_i - \mu_j)^2}{\sigma_j^2} - 1 \right] \right\} \quad (14.4)$$

Thus, as the means of the two distributions move apart, relative to the variances, the probability of a Type II error becomes increasingly closer to zero.

14.3.2.2 Performance Prediction in the Bayesian Framework

The Bayesian framework, in contrast, applies to problems in which the total probability of error (P_E) is more important than the probability of a Type I error. In this context, P_E bounds can be approximated via

$$P_E = \exp(-C_{i,j}) \quad (14.5)$$

where $C_{i,j}$ is the Chernoff information between targets i and j . This is given by

$$C_{i,j} = -\min_{0 \leq \lambda \leq 1} \{\mu(\lambda)\} \quad (14.6)$$

where

$$\mu(\lambda) = \ln \left[\int_x p_2(x)^\lambda p_1(x)^{1-\lambda} dx \right] \quad (14.7)$$

For two Gaussian distributions with different means and variances, this reduces to

$$\begin{aligned} \mu(\lambda) = & \frac{\lambda}{2} \ln \left(\frac{\sigma_1^2}{\sigma_2^2} \right) + \frac{\mu_1^2 \sigma_2^2 (\lambda - 1) - \mu_2^2 \sigma_1^2 \lambda}{2\sigma_1^2 \sigma_2^2} \\ & - \frac{(\mu_2^2 \sigma_1^2 \lambda - \mu_1^2 \sigma_2^2 (\lambda - 1))^2}{2\sigma_1^2 \sigma_2^2 (\sigma_2^2 (\lambda - 1) - \sigma_1^2 \lambda)} + \dots + \frac{1}{2} \ln \left[\frac{\sigma_2^2}{\sigma_1^2 \lambda - \sigma_2^2 (\lambda - 1)} \right] \end{aligned} \quad (14.8)$$

Again, the less the probability density functions (pdfs) overlap, the smaller the probability of error.

14.3.2.3 Comparison of Algorithms Via the Wald Sequential Test

The Wald test can be used to compare any two systems that produce binary responses, rendering it suitable to compare ATR systems in a pair-wise fashion [3]. By theoretical design, this test requires fewer samples than a confidence interval or ranking and selection approach to choose the superior system [3]. Furthermore, unlike other tests, it remains valid even when independent sampling from two processes cannot be performed [3].

The basic Wald test can be performed using the ratio of successes to failures for each process, or the efficiency (k) of the process, risk tolerances (α and β), and counting variables (t) [3]. The ratio of efficiencies between the first and second process is defined as u ; u_0 and u_1 are set such that errors yielding values of $u \leq u_0$ or $u \geq u_1$ are considered significant [3]. The counting variable, t , is equal to t_1 plus t_2 , where t_1 is the number of times an image is classified correctly by the first ATR process but not the second and t_2 is the corollary for the second ATR process [3]. Lower and upper bounds are defined by

$$\text{lower bound: } \frac{\log \frac{\beta}{1-\alpha}}{\log u_1 - \log u_0} + t \frac{\log \frac{1+u_1}{1+u_0}}{\log u_1 - \log u_0} \quad (14.9)$$

and

$$\text{upper bound: } \frac{\log \frac{1-\beta}{\alpha}}{\log u_1 - \log u_0} + t \frac{\log \frac{1+u_1}{1+u_0}}{\log u_1 - \log u_0} \quad (14.10)$$

If t_2 is below the lower bound for any value of t , then process 1 is better than 2 [3]. Similarly, if t_2 is above the upper bound, then process 2 is better [3]. If t_2 is between the bounds, testing continues [3].

14.3.3 Using Performance Prediction Tools in the Design Process

Performance prediction techniques have great utility in the design phase of an ATR scheme. To demonstrate this, consider the following list of questions likely to be considered when designing an ATR scheme.

- Have enough features been selected in the feature set?
- Is another feature going to improve overall results enough to justify the cost of collecting it?
- Can likely ATR errors be predicted to inform design of subsequent processing steps, hence improving overall system robustness?
- How much will variations in target articulation affect the robustness of the ATR answer?
- What are the performance ramifications if the features are corrupted to varying degrees?
- How long must the features be collected to achieve the desired level of reliability?

While high-fidelity models could be developed and exercised to address these questions, extensive high-fidelity modeling is not likely to be computationally efficient or cost effective. Performance prediction techniques, in contrast, can be leveraged to address these questions in a computationally efficient, cost effective manner. For example, to

characterize the degree of observation degradation tolerable to an ATR scheme, the ATR designer could run exhaustive sensitivity analysis of the algorithm. A more efficient approach would be to characterize the ramifications of corruption on the feature pdf and use this to efficiently update the predicted confusion matrices and probabilities of Type I and Type II errors. While the former approach would require both generation of corrupted feature sets and execution of the actual ATR algorithms, the latter approach requires only generation of corrupted feature sets and computation of either the Chernoff information or relative entropy. Hence, appropriate use of performance prediction techniques could enable the ATR designer to test a much broader range of conditions than could be addressed solely using prototype ATR code.

14.4 | SYNTHETIC APERTURE RADAR

ATR in radar systems is often performed via SAR images (see Chapters 6 and 7), which provide two-dimensional magnitude and phase data related to the target signature and thus provide valuable information for recognizing complex stationary targets. In fact, since SAR is the most common form of radar-based ATR, this section constitutes the majority of the chapter.

The Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset and its approach to ATR are well known and will be described briefly here as an introduction to some of the components of SAR ATR. The ATR methodology developed in the MSTAR program includes six modules: focus of attention, indexing, search, predict, extract, and match. The focus of attention (FOA) module searches a large image and extracts small image chips using a constant false alarm rate (CFAR) algorithm; these chips will be considered in future modules [4]. Subsequent sub-stages in the FOA then attempt to discard non-target chips by looking at global features, such as size, shape, contrast, and fill ratio, as well as using pattern recognition discrimination [4]. The indexer is the first stage of the classification sub-system, producing initial hypotheses for target type and orientation through zero-crossing and ridge features [4]. The predict, extract, match, and search (PEMS) modules serve as the extended classifier, if necessary [5]. The predictor module uses a computer-aided design (CAD) model of the target and Xpatch[®] Electro-magnetic Simulation Software (SAIC, McLean, VA, USA) to produce synthetic SAR data. The matching module compares the predictions with extracted features [4]. The extract module will take an image chip and instructions from the search module and output a set of features for search [6]. Classification is complete when all hypotheses given by the Indexer have been considered by the search module [4].

14.4.1 Step 1: Identify the Target Set

ATR for SAR requires the evaluation of numerous SAR images for training and testing algorithms across the target sets of interest. Several publically released SAR datasets are available for such purposes, including the MSTAR [7], Gotcha, Sandia National Laboratory and General Atomics Lynx sensor [8], Advanced Detection Technology Sensor (ADTS), and Georgia Tech Research Institute (GTRI) [9] datasets. The acquisition parameters and target sets are listed in Table 14-2.

In addition, SAR simulation software has been developed to generate synthetic SAR imagery including noise, speckle, layover (see Chapter 7 for details), foreshortening,

TABLE 14-2 ■ Publicly Released SAR Data Sets

	MSTAR	Gotcha	Lynx	ADTS	GTRI
Resolution	0.3 m × 0.3 m		1 ft.		1 ft × 1 ft
Polarization	HH	Full	Unknown	Full	Full
Depression angles	17°	8 angles	41–44°	22.5° and 40°	27.99° to 31.93° in 0.0138° steps
Aspect angles	360°	360°	360°	325–training angles, 68 test angles	0° to 360° in 0.05° steps
Frequency	X-band, 9.6 GHz	X-band	Ku-band	Ka-band, 32.6–37 GHz	X-band, 9.6 GHz
Target types	Military (armored personnel carriers, battle tanks, vehicles), confusors	Civilian vehicles and calibration targets	Military (armored personnel carriers, tanks)	Armor/military vehicles	Military (M2's and T72)
Environment types	Natural and man-made (cultural) clutter (~30 km ²)	Parking lot	Rural, suburban, and urban, NM and FL (10 km ²)	Stockbridge, NY (rural) and Ayer, MA (Urban) in stripmap mode	

corner reflector effect, and edge currents [10]. Because SAR imagery of a complex object is highly aspect dependent and affected by overlay at steep incidence angles [10], it is necessary to acquire images at many aspect and depression angles (see Figure 14-1 for an example of how target features change with depression angle) or to develop the requisite simulation capabilities. It should also be mentioned that while image libraries exist for microwave and millimeter-wave systems, libraries in ultra high frequencies (UHF) are limited [11].

14.4.2 Step 2: Select the Feature Set

The available features and their quality depend heavily upon the type of radar used to observe the targets. SAR radars vary widely with application, encompassing such diverse forms as polarimetric SAR, bistatic polarimetric SAR, millimeter wave SAR, and ultra-wideband SAR.

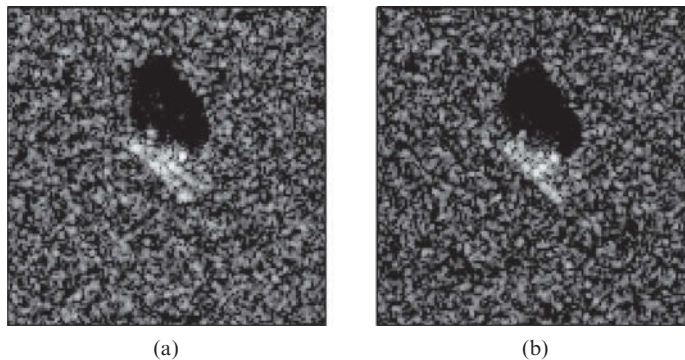


FIGURE 14-1 ■ BMP-2 vehicle at two depression angles:
a) 15 degrees and
b) 17 degrees.
(Reproduced from
MSTAR dataset with
permission.)

14.4.2.1 Types of Sar

Fully polarimetric SAR imagery has been shown to improve target-detection performance as well as the probability of correct classification. Polarimetric data can be optimally combined using the polarimetric whitening filter to minimize speckle, sharpen object edges, reduce clutter variance, and enhance target signatures relative to the background (see 14.4.3.2.2) [12]. Principal component analysis (PCA) can also be used to fuse four polarizations into one [13]. Using one target type on the GTRI dataset, Mishra and Mulgrew found that with similar elevation angles the data are less separable with multi-polar imagery than with single polarization; conversely, with a large span of elevation angles, multi-polar imagery increases the separability of the data [14].

Bistatic polarimetric SAR also appears to show promise. Based on limited simulations, its potential has been shown without considerable loss in performance compared with existing monostatic systems. Increased bandwidth can compensate for degradations in ATR performance caused by increasing bistatic angle. Bistatic polarimetric data contains additional information about the target and possibly its physical features, much like its monostatic counterpart [15].

Millimeter wave (mmW) SAR is appropriate in some applications that require high-resolution imaging, since new features for discrimination and classification emerge in this scheme. Objects that are point targets at lower resolutions can be resolved into extended targets, and the polarimetric characteristics of individual scattering centers should be more pronounced [16].

Ultra-wideband SAR is most helpful when targets are hidden in foliage, as it can detect them by identifying target resonance effects. Fourier and multi-resolution bases, such as the Haar wavelet and Gaussian basis, can be used to extract the spectral information necessary to detect wideband resonance effects [17].

14.4.2.2 Candidate Features

Depending on the application and type of SAR being used, a variety of features may be extracted. These features generally fall into the following categories: polarimetric features, texture features, contrast features, and size features.

Polarimetric Features Common polarimetric features used for finding objects that follow the even-bounce nature of some targets of interest, such as tanks and howitzers, include: percent pure (odd and even), percent pure even, and percent bright even [1]. Odd bounces are defined as $|HH + VV|^2/2$, while even bounces are $|HH - VV|^2/2 - |HV|^2 - |VH|^2$, where HH, VV, HV, and VH are component intensities [18]. A pure odd bounce pixel is one where the intensity of the odd component is greater than 75% of the total intensity and similarly for pure even bounce pixels [18]. Other geometric and polarimetric features include the range extent and Doppler extent of the 10 strongest VV scatterers [16].

Some features are based on the normalized Stokes vectors, $\vec{S} = (S_1, S_2, S_3) = \left(\frac{Q}{I}, \frac{U}{I}, \frac{V}{I}\right)$ where I, Q, U , and V are the Stokes parameters, and the ellipticity, ε , of the polarization ellipse that describes the angular distance between the equator and the polarization state on the Poincaré sphere (graphical tool for visualizing different polarizations). Specific polarimetric features are: the fraction of the 10 strongest VV scatterers where the product $\mathcal{P}_{HV} = \vec{S}_H \vec{S}_V$ is < -0.9 , and the fraction of the 10 strongest VV scatterers where $|\varepsilon_H|$ and $|\varepsilon_V| < 10^\circ$ [16].

Texture Features Common textural features used for discrimination of man-made targets from natural clutter include: standard deviation, fractal dimension, and weighted fill [1]. Standard deviation is simply the standard deviation of the data within a target-sized template to measure the fluctuation of pixel intensities, which are generally larger for targets than natural clutter [12,19]. Fractal dimension measures the spatial distribution of the brightest scatterers in the detected object and can be calculated as

$$\dim(S) = \lim_{n \rightarrow 0} \frac{\log M_n}{\log \frac{1}{n}} \quad (14.11)$$

where M_n is the minimum number of $n \times n$ boxes needed to cover the detected object, S [12]. The weighted-rank fill ratio measures the fraction of the total power present in the brightest 5% of the detected object scatterers, which is higher for targets than natural clutter [12].

Contrast Features Common contrast features used for rejection of objects without a sufficient local brightness include: peak CFAR, mean CFAR, and percent bright CFAR [1]. To be more specific, peak CFAR is the maximum value within the target-shaped blob in the CFAR image created by the algorithm described in 14.4.3.3.1 [1]. The mean CFAR feature is the average CFAR over the target shaped blob, and the percent bright CFAR is the percentage of pixels within the target-shaped blob that exceed a threshold set based on the training data [1].

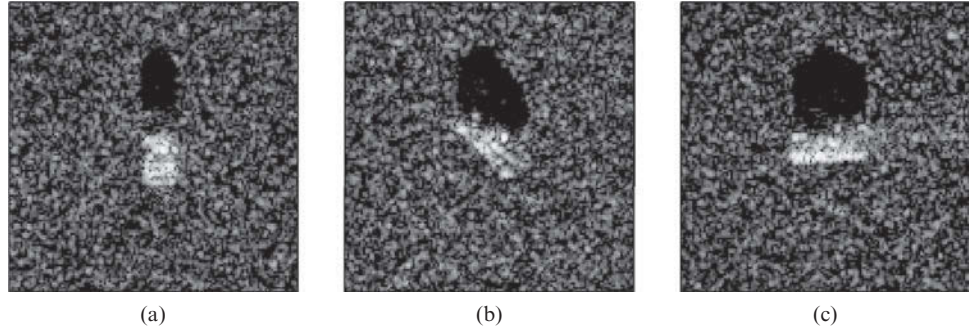
Size Features Common size features used for rejection of inappropriately sized objects compared with the targets of interest include: mass, diameter, and normalized rotational inertia [1]. The Radon transform can be used to find the length and width of a target. First the target chips from the pre-screener are segmented into background, shadow, and target classes using two-parameter CFAR, morphological processing, and feature clustering. The 2D Radon transform is then applied to the binary image of just the target. The height, or maximum Radon transform value, for each projection angle is found. The width (W) is found as the projection angle with the minimum height, while the diagonal (D) has the maximum height. The length is $\sqrt{D^2 - W^2}$ [19].

Additional Features To reduce false alarms from building returns, features from different types of SAR imagery can be exploited to identify and ignore buildings. In single strip-map SAR images, cardinal streaks and supporting shadows can be used to identify buildings. Multi-pass airborne SAR images can be registered to extract 3-D heights and linear patterns associated with buildings. Finally, local histogram thresholding can be used to segment buildings from the ground in interferometric SAR images [20].

Most ATR performed on SAR images is done to identify stationary targets; however, some features can be extracted for classifying moving targets. For instance, high range resolution (HRR) profiles can be extracted from SAR images as motion-invariant features for identifying moving ground targets; HRR profiles formed in this way have a higher target-to-clutter signal ratio than those formed from raw data because much of the ground clutter is removed during Doppler processing in the process of forming a SAR image [21].

Feature Dependency on Angle Because the aspect angle to an unknown target is likely unknown itself, the discrimination phase of ATR may begin with an algorithm to estimate

FIGURE 14-2 ■
BMP-2 vehicle at
three aspect angles:
a) 0 degrees,
b) 45 degrees, and
c) 90 degrees.
(Reproduced from
MSTAR dataset with
permission.)



the position and orientation of the candidate target (see Figure 14-2 for an example of changing target features with aspect angle). Pose estimation is useful in that it (1) is required for simulating complicated physical mechanisms for developing a set of model signatures against which to compare test images, (2) optimizes the search time by reducing the model database that must be searched, and (3) allows more features to be exploited in the recognition task [22]. One algorithm for pose estimation translates and rotates a target-sized rectangular template around the image until the energy within the template is maximized [1]. A more complex, but robust, scheme is to find closed-contour signatures of the target through segmentation and then apply four independent, but ranked, criteria for finding the aspect angle. Three of these criteria are determined from the information provided within the smallest rectangle, or bounding box, completely encapsulating the target: target-to-background ratio (TBR), perimeter, and edge pixel count. The aspect angle yielding a rectangle with the maximum TBR within the box, minimum perimeter of the box itself, and maximizing the number of pixels on the target contour that are in a close neighborhood to one of the major edges of the box [22]. To counter the limitations of the bounding box criteria caused by box misalignment from anomalies in the segmentation process, the Hough transform criteria detects straight lines in the image and the orientations of the longest three straight lines in the image of the target contour are used to approximate the target pose [22]. Although these criteria work well, the cardinal points of 0° and 180° present a generic, round-shaped signature that may require an additional procedure to identify [22].

The Hilbert-Schmidt (minimum mean-squared error) estimator is yet another method of target aspect estimation [23]. For objects on a flat surface, target orientations exist in the special orthogonal group of dimension 2 and can be represented by a 2×2 matrix with determinant one [23]. The Hilbert-Schmidt distance between two rotation matrices,

$A = \begin{bmatrix} \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_1 & \cos \theta_1 \end{bmatrix}$ and $B = \begin{bmatrix} \cos \theta_2 & -\sin \theta_2 \\ \sin \theta_2 & \cos \theta_2 \end{bmatrix}$, can then be expressed as

$$\begin{aligned} d_{HS}^2(A, B) &= \|A - B\|_{HS}^2 = \text{tr}[(A - B)^T(A - B)] = 2n - 2\text{tr}[A^T B] \\ &= 2n - 2\text{tr}[B^T A] = 4 - 4\cos(\theta_1 - \theta_2) \end{aligned} \quad (14.12)$$

where n is the dimension of the rotation matrices (i.e., 2) [23]. In this way, orientation can be estimated by finding the orientation that minimizes the conditional mean (conditional on the SAR image and target type) of the squared Hilbert-Schmidt distance (see [23] for more details).

14.4.3 Observe the Feature Set

A great deal of pre-processing is typically necessary to extract the desired features from a SAR image. This sub-section discusses some of the major pre-processing tasks typically associated with SAR ATR.

14.4.3.1 Observation Conditions

Images acquired with certain radar acquisition parameters have been shown to yield improved ATR performance. The probability of correct classification improves with finer resolution imagery [7], as the resolution must be equal to or finer than the mean local distances between neighboring persistent scatterers for the best results [24]. In addition to resolution, SAR imagery is affected by depression angle, squint angle, frequency, PRF, polarization, single/multi-look, sensor abnormalities, noise level, and stripmap versus spotlight mode [25]. SAR ATR performs close to perfect when testing and training are near the same conditions [26].

Different operating conditions between the training and testing datasets can lead to poor ATR performance. In addition to the radar acquisition parameters listed in Table 14-1, minor differences in target configuration can make a dramatic impact. For instance, the presence or absence of barrels on tanks [7] is an influential configuration change. This intra-class variability can decrease interclass separability and degrade ATR performance [27]. The classification performance of targets in revetments is also sub-standard due to obscuration [27], where performance worsens as more of the target is obscured [28]. Natural camouflage (e.g., covering with branches), camouflage by radar nets, and confusing by adding scattering centers from multi-path effects all have an impact on ATR performance [28]. On the other hand, decoys designed for optical and infrared applications appear to have no impact on SAR ATR performance [28]. While most of the aforementioned extended operating conditions degrade the probability of correct identification, or classification, they have little to no impact on the ability to simply detect the presence of a possible target from clutter, or discrimination [26].

14.4.3.2 Image Formation and Pre-processing Techniques

Superresolution High-definition vector imaging is a method of producing higher resolution SAR imagery with reduced sidelobes and clutter [29]. This method uses two-dimensional minimum-variance techniques based on the Capon algorithm (also known as maximum-likelihood method, minimum-variance distortionless response, and reduced-variance distortionless response) and a two-dimensional version of the multiple signal classification (MUSIC) algorithm [29]. Capon's algorithm is used to estimate the RCS of the scatterer, or power, and image, or power spectrum [29]. The MUSIC algorithm is used to estimate how closely the data matches a scattering model [29]. MIT Lincoln Laboratory has shown that ATR performance improves significantly when high-definition imaging is applied to low-resolution imagery (e.g., 45% to 73% correct classification with $1.0 \text{ m} \times 1.0 \text{ m}$ to $0.5 \text{ m} \times 0.5 \text{ m}$ improved resolution) [7].

Polarimetric whitening filter An optimal polarimetric processing technique to combine HH, HV, and VV polarization components is the polarimetric whitening filter (PWF) [1]. This processing enhances image quality by minimizing speckle and sharpening object edges leading to improved performance in each stage of ATR [1]. Compared with HH

data, PWF has been shown to result in a reduction in the number of false alarms at the output of the pre-screener, an improvement in the rejection of man-made discretely during discrimination, and a reduction in PFA for a given PD after completion of the entire ATR system [1].

Frequency domain analysis with complex image Sensitivity to collection differences between test images and target chips is important to address in ATR processing. The location of the test target in the imaging scene as well as the flight path of the radar platform can lead to spatial warping between test images and target chips. Slant plane variations and erroneous shifts can also arise depending on the SAR imaging algorithm used (see Chapter 7). Finally, sensor variations can result in different 2D image point responses (IPR) or point spread functions between the reference and test SAR images [30]. One alternative solution is to perform ATR on the SAR phase history data instead, thereby eliminating sensitivity to these problems. Using signal subspace matched filtering, the complex SAR signature of a target region of interest can be compared with a reference signature as a function of radar frequency and aspect angle [31], using 2D adaptive filtering methods to calibrate and compensate for the variations of the IPRs [30].

14.4.3.3 Pre-screening and Discrimination Algorithms

CFAR A CFAR detector can be used as a prescreener for selection of candidate targets based on local brightness. An example CFAR algorithm for these purposes requires that the peak target-to-average clutter ratio (in dB) over the log standard deviation of the clutter be above a constant controlling the false alarm rate to be considered a candidate target pixel [1]. Clutter estimates are often obtained outside of a guard area around the potential target, which is defined using the target type and image resolution [1]. Other CFAR prescreeners have been proposed, such as a two-stage CFAR detector. In the first stage, CFAR detects possible targets while global statistical parameters are used to remove false targets. Then, more targets are detected in regions adjacent to those detected in the first stage using local statistical parameters. Finally, conditional dilation (an image morphological operation) is used to recover target pixels lost in the second stage [32].

Another pre-screening algorithm with three components was developed by Georgia Tech. In the first component, a binary image is formed, consisting of the pixels that occur in a high density of bright returns and are not in shadow, defined as having high brightness compared with surrounding areas. The second component removes isolated bright returns using a rank order filter. The last component rejects clutter that is not within the bounds of the expected target size [33].

Vector Quantization Vector quantization is a data compression technique that can be used to combine features into a discrimination statistic. Initial decision surfaces in feature space are specified by a code book. This is constructed by setting the code vector equal to the k -th input training vector, where k is $\sum_n \left| \frac{X_j(n) - Y_i^l(n)}{Y_i^l(n)} \right|$ minimized over i and maximized over j in a generalized Lloyd (LBG) design framework, X is the training data, Y is the code-vector, and L is the number of code-vectors. The number of code vectors in the code book used for all targets is increased until every cluster is nearly homogeneous. During testing, vector quantization uses a maximum likelihood criterion to search for the most likely feature over all possible feature vector models [33].

Universal Classifier The universal classifier can be used to find optimal asymptotic solutions to non-Gaussian properties of signals, such as in the identification of target-like objects in a scene [34]. This algorithm uses statistical distance measures between the test image and training library without assuming a Gaussian environment [34]. Empirical probability density estimates can be formed by quantizing the training data. Then, the Kullback-Leibler distance between the unknown probability density of the target under test and the densities in the library is minimized to find the most likely hypothesis – a target is present or not [34]. This method performs well with partially obstructed or occluded targets [34].

De-noising To aid in the ATR process, de-noising of the imagery may be performed in a pre-processing step. This can be accomplished through the use of the stationary discrete wavelet transform [35]. This will filter the signal into a high and low-frequency part, where the high-frequency signal contains mostly noise and can be discarded [35]. However, some high-frequency details in the image, like the target shadow or the edges of roads, are also contained in this component; thus, a scale correlation method can be used to extract and retain this information [35].

Cluster Discrimination A subset of calculated discrimination features can be combined into a single discrimination statistic as a quadratic distance metric,

$$\vec{d} = (\vec{h}^T \mathbf{R} \vec{h})/n \quad (14.13)$$

where n is the number of features in the subset, \vec{h} is the vector of features measured from the detected object minus the mean value of the features from training data, and \mathbf{R} is the covariance matrix of the features in training data. A large quadratic distance means the candidate target is likely a clutter false alarm and should be rejected. The best subset of features to use varies with polarization and resolution [1].

Segmentation Effective segmentation of the target, background, and shadow can aid in successful discrimination. One method of accomplishing this is to first model the image gradation as a generalized mixed Gaussian distribution by estimating the parameters with the expectation maximization method [35]. Anisotropic physical heat diffusion can then be used on the posterior probability with respect to the pixel gradation to remove multiplicative noise and smooth the middle of each region while preserving the edges [35].

Fourier Transform For neural network classifiers, the Fourier coefficients from application of a fast Fourier transform are used as features. The image is first normalized by subtracting off the mean and dividing by the standard deviation. Then, the real and imaginary Fourier coefficients are found and mapped to a polar angle between 0 and π . The most invariant coefficients, or those with the longest independent polar angle vector lengths, are then defined as features [36].

Multi-Resolution Analysis Multi-resolution analysis (MRA) exploits the automatic feature isolating properties of the wavelet transform [34]. The wavelet transform gives spatial and frequency localization by decomposing the image into four subimages at each level of decomposition: the low-frequency components in horizontal and vertical

directions (LL); low-frequency components in horizontal and high frequency in vertical (LH); high frequency in horizontal and low frequency in vertical (HL); and high frequency in both horizontal and vertical (HH) [37]. The image is expressed in terms of the sum of bases functions that are shifted and scaled versions of a mother wavelet; the mother wavelet (waveform of certain shape and duration) that minimizes intra-target variability and maximizes inter-target variability is deemed the best [37]. Wavelet decomposition using a reverse biorthogonal wavelet pairs mother wavelet yields good ATR performance [37].

Principal Component Analysis Principal component analysis (PCA) can be used to extract individual features, or eigenvectors, in an image and recombine only those containing target information. As a result, a SAR image can be reconstructed with little noise and the removal of similarities and redundancies that add no extra information for the classifier [2]. By optimizing the principal component number (number of eigenvalues considered significant) to have an accumulative contribution rate appropriate for the quality of the image, the target recognition rate can be improved [38]. The computational complexity of PCA can be reduced by using 2D PCA, where the covariance matrix, from which the eigenvectors are extracted, is computed using the maximum likelihood method on a set of 2D images [39]. This method only eliminates the correlations between rows; therefore, a two-stage implementation of 2D PCA can be employed to find the column projection matrix and further reduce the feature space [40]. The features can be used as inputs to support vector machines (see section 14.4.4.16) for classification [39].

As an aside, Fisher's linear discriminant analysis (LDA) has been reported to be a better feature extractor than PCA for many applications. However, one study looked at the performance of LDA compared with PCA for SAR ATR and found that PCA produces better classification results with better confidence than LDA [13]. The eigenvectors, or features, in LDA are found using not only the within class covariance, like PCA, but also the between class covariance [13]. The performance of LDA may improve with larger training datasets [13].

Independent Component Analysis Independent component analysis (ICA) is a statistical method that can extract non-Gaussian, mutually independent components, or features, from observed data without prior knowledge of how they were mixed [41]. This method can serve as a more generalized alternative to PCA. As with PCA, these features can be used as inputs to support vector machines (see section 14.4.4.16) for classification [41].

14.4.4 Test the Feature Set

14.4.4.1 Statistical Training Libraries

A SAR image can be thought of as a complex Gaussian random process with a conditional mean, covariance matrix, and complex Gaussian noise. SAR data can be well modeled with zero mean and a diagonal covariance matrix [23]. Model training involves estimating the variances of each pixel, which can then be used for pose estimation and target recognition [23]. Target classification can be accomplished using the estimated parameters of the Gaussian process and a Bayesian approach, where one sums over the complete set of unknown target poses and positions [23]. Recognition rates of over 97% for a ten class library with standard operating conditions can be achieved with this method [23]. More on the Bayesian approach is discussed in 14.4.4.13.

14.4.4.2 Template Libraries

In a template-based approach to ATR, templates are created from collected SAR imagery and compared to the test data. The template can be a spatial signature for one specific target geometry or state, or it can be a linear combination of registered signatures over several conditions [42]. One example of combining signatures over several geometries is the MIT Lincoln Laboratories ATR approach in which target image templates were generated by averaging five images acquired over 5° of aspect angle [7].

14.4.4.3 Adaptive Feedback

Rather than using training data based on supervised learning or solely on *a priori* image collections, a method for using feature-based classification in conjunction with truth feedback to improve ATR learning and classification over the course of a mission can be employed [43]. By doing so, operational conditions that are not included in the training data can be addressed. In the Air Force Research Laboratory's (AFRL) AdaptSAPS system and Sensor and Data Management System (SDMS), truth feedback is assumed to come from human analysts with access to other information sources [43]. The quadratic distance measure determines which class an object belongs to and changes adaptively by refining the feature vector mean and covariance as more data is received during the mission [43].

14.4.4.4 Models

Model-based ATR generates features from mathematical models of the sensing process and target geometry, which can then be matched to features from unknown signatures [42]. One well-documented model-based algorithm is the PEMS modules of the MSTAR program, which follows the hypothesis-and-test approach for hypothesizing a target type, extracting the appropriate feature set from models, performing matching, and determining the likelihood score [42]. This iterative ATR subsystem uses high-fidelity models of targets to refine and verify target classification [4], as described earlier.

The European Aeronautic Defence and Space (EADS) Company has created a radar image generator called SAR-VIEW which consists of a terrain database, feature map construction, height map construction, radiometric expectation values, CAD models for targets, and formation of a synthetic SAR image. First, a set of terrain data is extracted from the terrain database. Using pre-defined requirements for the resolution of the image and knowledge about the platform heading, a landscape module transforms and interpolates the terrain for the desired conditions to produce a map of height values and a map of land cover types. The expected clutter RCS at each pixel of the map can then be calculated from ground and slant ranges. A combination of geometric and physical optics is used to find radar cross sections from 3D CAD models of man-made targets. After the targets are integrated with the landscape, pixel correlation and speckle effects are added to give the final synthetic SAR image. Additional approaches for improving the image realism can be employed in this system. These images are useful for ATR training and for integration into combat flight simulators [44].

14.4.4.5 Persistent Scatterers

Persistent scatterers have been shown to exist over nearly the entire range of possible target orientations in real and synthesized (XPatch) SAR images [45]. They can be extracted from higher resolution SAR images or from Xpatch electromagnetic signature predictions of 3-D CAD models [46].

Scattering centers can account for a large percentage of the overall energy in an image of a target [47]. The scattering center attributes from synthetic SAR images of targets that are larger than a few wavelengths can be determined using the geometric theory of diffraction model [47]. Attributes of interest include location, fully polarimetric amplitudes, aspect-dependent amplitude tapers and geometric type [47]. These features should be augmented by other features, such as texture features or shadow, when used for ATR [47].

14.4.4.6 Evolutionary Programming

Evolutionary programming can be used as a model optimization technique in ATR through a defined number of mutation phases on the parameters of interest. For example, an object may be represented as a small set of spatially-localized filters. Evolutionary programming can be used to generate and to optimize spatial filter locations and coefficient values for the purpose of classifying targets using intra-region pixel relationships [48]. In other words, the evolutionary programming algorithm can optimize the locations and coefficient values to approach the largest target-clutter separation distance.

14.4.4.7 Hybrid (Image Library and Model)

A typical template library is quite large and must be coarsely sampled to span the entire hypothesis space. One method of reducing the total number of templates required for handling various collection geometries, obscuration, and articulation conditions and increasing the sampling resolution of the hypothesis space is to use a hybrid model/template approach. The library size can be reduced by using a model to determine the transformation or perturbations necessary to adapt an existing template to the signature of a neighboring sensor geometry or target articulation, while maintaining some of the computational advantage of the pure template-based approach [42]. An alternative hybrid approach for increasing the resolution over the hypothesis space can be constructed by using the templates to reduce the number of hypotheses that the models must search over; however, a large template set is still required for this approach [42].

14.4.4.8 Mean-Squared Error (MSE)

The first algorithm to be discussed for classifying targets, the mean-squared error algorithm, is a template-based approach [49]. This algorithm compares the image under test with a pre-defined, normalized, and windowed set of template images for target types of interest. The normalization process should remove the clutter background level to eliminate errors from absolute RCS calibration of the sensor data, while windowing eliminates the influence of nearby clutter in the classifier. The MSE is calculated over the pixels in the reference template window to calculate the relative difference in total power. If the MSE between the test image and the templates for the given target types always exceeds a maximum allowable threshold, then the detected object is defined as clutter; otherwise, the object is classified under the target type corresponding to the lowest MSE [1]. With this approach, the number of target classes can be doubled with only a slight decrease in ATR performance [27].

14.4.4.9 Distances Between Features

The Hausdorff distance is tolerant of small position errors [50] and is invariant to rotation given scale and signal level normalizations have been performed [51]. The directed Hausdorff distance finds the point in the first set that is farthest from any point in the

second set and then measures the distance from that point to the nearest neighbor in the second set. The general Hausdorff distance is the maximum value of the two directed Hausdorff distances between two sets [51]. The average Hausdorff distance is resilient to stray pixels and, thus, is useful for SAR classification [52]. Several modified Hausdorff distance measures exist to improve upon reliability and robustness [51].

The Pearson distance has been shown to perform better with Fourier descriptors (see 4.3.1.1) than the Euclidean or Mahalanobis (scale-invariant and takes into account data correlation) distances. The Pearson distance is given by $d(p,q) = \sqrt{1 - C(p,q)^2}$, where $C(p,q)$ is the normalized correlation coefficient between the test and training Fourier descriptors, p indexes over target class, and q indexes over aspect angle [53].

14.4.4.10 Invariant Histograms

A histogram of weakly invariant feature values, such as points and line segment features, can be used to classify targets [54]. Pairs of features that are invariant with translation are included in the histogram. Distance measures can be computed between the invariant maps and the candidate images [54].

14.4.4.11 Likelihood-Based Classifier

By assuming the SAR image is a complex Gaussian vector, the generalized likelihood test can be used to find the target type and orientation that maximizes the logarithm of the likelihood function of the test image [46]. The mean of the test image is assumed to be zero due to the random phases in SAR imagery, and the pixels are assumed to be statistically independent, yielding a diagonal covariance matrix [23]. Thus, the log-likelihood function can be simplified to:

$$l(\hat{f}|\hat{\theta},a) \propto \sum_i \left[-\log \mathbf{K}_{i,i}(\hat{\theta},a) - \frac{|\hat{f}_i|^2}{\mathbf{K}_{i,i}(\hat{\theta},a)} \right] \quad (14.14)$$

where the i subscripts denote the i -th element of a vector or diagonal matrix, \hat{f} is the observed image, $\hat{\theta}$ is the estimated pose angle, a is the target type, and $\mathbf{K}_{i,i}(\hat{\theta},a)$ is a matrix of the model variances estimated from the training data by

$$\mathbf{K}_{i,i}(\hat{\theta},a) = \frac{1}{N_k} \sum_{\theta \in W_k} |(\hat{f}^{r_k}(\theta,a))_i|^2 \quad (14.15)$$

where $(\hat{f}^{r_k}(\theta,a))_i$ is the i -th pixel of an aligned training image, W_k is the range of angular orientations used for training θ , and N_k is the number of training images within that range [46,23].

14.4.4.12 Neural Nets

Pose estimation is an important component in the process of classification. By estimating the aspect angle, a smaller library of template images needs to be compared to the target chip under test. While simple methods like rectangular fitting [1] and the Hough transform [22] can be used to estimate the target aspect, more sophisticated methods such as multilayer perceptrons and linear networks have shown average estimation errors of less than 10 degrees [55].

In addition to pose estimation, multilayer perceptron neural networks can be used for classification based on extracted features. These networks consist of an input and output layer in contact with the outside world and a variable number of hidden layers

depending on the application [36]. Because the complexity of the network and the number of neurons is closely related to the length of the feature vector, care must be taken to optimize and minimize the number of features used for classification to save computational resources [56]. The multilayer perceptron, as well as other classification neural networks in the literature, use the back-propagation training algorithm [56–57]. The procedure for this training method is as follows: (1) the weights are initialized to small values; (2) the output of the error is calculated; (3) the partial derivative of the error is calculated with respect to each weight; (4) the weights are updated using the previous weights, the partial derivative, and the preset learning rate value (how far to move the weights in the direction of the partial derivative for each iteration); and (5) the process is repeated until the training set is exhausted or the output error of the network reaches an acceptable level [36]. Classification is performed by initializing the network weights to those of the trained weights, inputting a sample from the testing set, and classifying the target based on the probabilities at each output node, or each target in the training set [36].

14.4.4.13 Bayesian Network

A Bayesian network is an acyclic directed graph composed of connected layers of nodes. In a Bayesian network, the nodes correspond to variables of interest, where the states can be discrete or continuous, while the arcs connecting nodes are generally given by conditional probabilities. Once such a network has been set up to model a problem, an updated classification of the target of interest can be made through posterior distributions of the state given the observed information [34]. The posterior distributions can be determined by Bayes' rule as $P(s|x) = p(x|s)P(s)/p(x)$, where $P(s|x)$ is a posterior probability, $p(x|s)$ is the state conditional probability, $P(s)$ is the a priori state probability, and $p(x)$ is a normalization factor equal to the summation of $p(x|s)P(s)$ across all states, s [21]. Such a network can aid in correctly fusing interdependent features for target classification [34, 52] or the hypotheses from individual features can be reported [58]. The Dempster-Shaffer algorithm may be an alternative to a Bayesian network when knowledge of priors and conditionals are not known in advance [52].

14.4.4.14 Nearest Neighbor

The nearest neighbor classification method computes the Euclidean distance between an unknown input vector and the mean vector for each target class. The target class with the shortest Euclidean distance is then identified as the target [21].

14.4.4.15 Hidden Markov Model

A hidden Markov model (HMM) is a probabilistic function of a Markov process [59] and consists of hidden states (S) and observable states (V), or measured features [60]. The HMM is defined by a state transition matrix (probability of going from one state to another), an emission matrix (observing a symbol when the system is in a certain state), and state priors. A suitable model is determined by attempting to maximize the likelihood of the model over the training data through expectation maximization [59–60]. During testing, multiple models can exist for each target. The probability of observing the test image is calculated for each of the models in order to classify the target chip [60]. Instead of images, features that are rotationally and shift invariant can be fed into the HMMs for training and testing; this approach can yield recognition rates as high as 99% [59]. The disadvantages of HMMs are that they do not capture any higher-order correlations between observations and

they assume the observation parameter distributions approximate a mixture of Gaussian or autoregressive densities closely [59].

14.4.4.16 Support Vector Machine

Support vector machines (SVMs) have been shown to have strong classification, fault tolerance, and generalization abilities [61]. SVM is a margin-based classifier that uses a subset of training samples to find support vectors giving the maximum decision boundary hyperplane of the classifier [62, 41]. For an SVM to discriminate between more than two classes, a parallel, cascading, or hierarchical decision scheme is generally employed [63]. SVMs cannot select an optimal feature subset from the complete extracted feature set; an optimal feature subset simplifies classification and can improve classification performance [61]. By combining an SVM with a rough set, or a rough estimate of the feature set consisting of lower and upper approximations, the classification phase of ATR can be performed well [61].

14.4.4.17 Correlation Filter

Correlation filters are used in the classification stage of SAR ATR because they are shift invariant, offering varying degrees of distortion tolerance, and are not tailored to a particular sensor or image properties [64]. Other advantages of advanced correlation techniques include: filter synthesis can be performed off-line; image preprocessing is unnecessary because of the filter's distortion tolerance and robustness; the processing complexity depends on the scene size but not its content; design optimization and performance bounding is relatively easy; and the algorithm maps well into hardware [64]. We discuss only a few of the many examples of these filters below.

14.4.4.18 Maximum Average Correlation Height

The maximum average correlation height (MACH) filter is an attractive correlation filter option because of its better distortion tolerance than other alternatives; however, it has poor clutter rejection performance because it relies strongly on the average training image. Variations of the extended MACH filter can be used to improve the clutter rejection performance [65]. The MACH filter generates peaks from which a peak-to-sidelobe ratio (PSR) is calculated; the test image belongs to the class with the highest PSR [64]. A MACH filter can be used in conjunction with a distance classifier correlation filter (DCCF) to improve performance by applying the DCCF if the MACH PSR is above an acceptable threshold [66].

14.4.4.19 Distance Classifier Correlation Filter

The distance classifier correlation filter (DCCF) classifies targets using the distance of the test image correlation to the class average correlation, instead of only the output correlation peak, and has shown better classification results than other correlation filters [66]. For best performance, interclass distance should be maximized and the average correlation energy minimized, while the average intraclass distance, represented by the average similarity measure, should be minimized [66]. The DCCF involves processing by a transformation filter to maximize the average spectral separation between classes and, thus, enhance discrimination; then, the shift-invariant minimum mean square error is calculated to determine the distance between the transformed test image and each transformed reference image [66]. The image is classified according to the class of the reference image yielding the smallest distance [66].

14.4.5 Summary of Synthetic Aperture Radar

The steps in the unified framework for ATR using synthetic aperture radar imagery have been discussed. Samplings of available datasets, feature types, pre-screening techniques, discrimination algorithms, and classification approaches found in the literature have been presented with brief summaries. This synopsis is meant to provide the reader with an appreciation for the breadth of techniques present in the SAR ATR literature and an understanding of how various algorithms fit within the Unified Framework.

14.5 | INVERSE SYNTHETIC APERTURE RADAR

Many of the ATR challenges and algorithms associated with SAR imagery are similar to those in ISAR imagery. Hence, rather than go through all four steps again (as in Section 14.4), we highlight the issues and techniques that are either unique to ISAR images or of particular concern when the target itself is moving, as in the ISAR imaging case.

14.5.1 Image Formation

This section describes a few challenges and alternatives in the ISAR image formation process that can make a significant impact on the performance of ATR.

14.5.1.1 Motion Compensation

Motion compensation for small water craft with more than one type of rotation (roll, pitch, or yaw) is particularly difficult [67]. One algorithm for this process involves establishing and tracking an image central reference point in range and Doppler and determining the rotation around the central reference point by tracking a secondary point [67]. Another algorithm uses Doppler tracking to remove higher frequency line-of-sight phase errors and stabilize the Range Track Doppler beam sharpening filters, range tracking to remove the line of sight range motion through range gates, and autofocus to remove low-frequency quadratic phase error [67]. Without adequate motion compensation, the resultant image may be improperly scaled or ill-focused [67], both of which present challenges to an ATR algorithm.

14.5.1.2 Estimating Angular Velocity and Acceleration

By assuming a rigid object and a low grazing angle, such that the line of sight is assumed to be in the horizontal plane, the target will exhibit angular motion about the centroid. Using this angular velocity, the instantaneous velocity in range can be calculated as the cross product of the distance between the point scatterers that make up the target and its centroid with the angular velocity projected into the range direction. The acceleration of the target is then simply the derivative of the instantaneous velocity. By converting the expression for acceleration into the coordinate system of the ISAR focus plane (range, cross-range, and displacement perpendicular to the slant plane but unobservable in ISAR) and applying the scale factor to convert cross-range to Doppler frequency, the partial derivatives of radial acceleration with respect to range and frequency can be expressed in terms of physical target parameters [68]. If these physical quantities are combined with estimation techniques, such as Kalman filters, the global angular velocity and acceleration of a rigid target can be estimated [68].

For ships, the yaw rate can be calculated from the Doppler difference between the bow and the stern, and the length of the ship. Similarly, an effective roll and pitch can be

calculating using the Doppler extent of the height features beyond the edge of the deck. Multiple image frames are necessary to calculate the real roll and pitch of a ship [69].

14.5.1.3 Using High Range Resolution Profiles

Much of the formation and utility of HRR profiles will be discussed later in this chapter, but this section will briefly describe their extraction from SAR images. HRR profiles become useful with ground target identification when irregular motion of a maneuvering target distorts SAR imagery. While HRR profiles processed directly from raw radar data will include an abundance of ground clutter, Doppler processing in the SAR image formation process removes much of the ground clutter from HRR profiles extracted along the down-range axis of SAR images. The down-range axis can be determined by the shadow orientation in the image. A cross-correlation template matching technique can be applied to classify targets based on HRR profiles; the performance of this method can be improved by fusing data from multiple looks after classifying each individually [21].

14.5.1.3 Using Different Frequency Subspaces

The bandwidth and center frequency of pulses influence the ATR performance on the resulting ISAR imagery. Linear chirped pulses with equal bandwidths but different center frequencies yield different recognition rates. However, information from chirps at different center frequencies can be combined to give a higher ATR response than the full bandwidth reconstruction with less processing and target movement knowledge required [2].

14.5.1.4 Time Frequency Processing

The short-time Fourier transform or the time-frequency distribution series can be used to form continuous range-Doppler images from overlapped ISAR images [67]. This technique allows high-resolution images to be formed at different times so that the rotation of the image through time can be monitored [67]. Radar frequency – fast time images can also be formed using the fast Gabor transform. This technique yields several other advantages, including: better target dimension estimation accuracy, better target rotation estimation, and the possible development of additional features in Doppler-time or radar frequency – fast time images [67].

14.5.2 Scattering Models

Statistically independent noise fluctuations exist between images of a moving target in ISAR imagery [70]. Additionally, the low rotational speed and weight of the target can lead to irregular sampling in azimuth and, thus, speckle noise and ambiguities [24]. To deal with these error sources, scattering centers are often extracted from ISAR images and used in subsequent recognition steps.

14.5.2.1 Filtering and Phase Analysis

Scattering centers can be extracted from ISAR and SAR imagery using filtering operations. For example, spatial variant apodization (SVA) seeks to achieve maximum spatial resolution while reducing sidelobes. By applying this algorithm to an ISAR image followed by phase analysis to extract the most robust scattering centers, a binary image of the dominant scatterers in a target can be constructed [24].

14.5.2.2 Pol-CLEAN

Another method for extracting scattering centers from polarimetric ISAR imagery is to use the Pol-CLEAN technique [71]. This algorithm finds the coordinates (delay-Doppler domain) of the brightest scatterer in the polarimetric ISAR image of interest, estimates target motion parameters to find the point spread function on the scatterer, and then removes the scatterer. This process is iterated for the next brightest scatterer until the energy of the image is below a given threshold [56]. The Pol-CLEAN technique can be followed by Cameron's decomposition to reduce the feature space dimension (algorithm outlined in [56]).

14.5.2.3 Euler Decomposition

In a polarimetric radar system, the Sinclair matrix, S , links the received, r , horizontal and vertical polarization fields to those transmitted, t , with a wave propagation term, k , as $r = kSt$. By forming a polarimetric ISAR image with multiple aspect angles and frequencies, a spatial map of an object's scattering properties can be formed and, thus, the Sinclair matrix for each pixel is known. However, the Sinclair matrix is not in a very intuitive phenomenological form, so it is advantageous to apply an Euler decomposition for the purpose of extracting the object's Euler scattering properties. Images of the five Euler parameters— m (max magnitude in dBsm), ψ (orientation in degrees), τ (symmetry in degrees), ν (bounce angle in degrees), and γ (polarizability in degrees)—can then be obtained [72]. A comparison of these test Euler images with those in a training library can then be used to classify the object under test, as described previously.

14.5.2.4 3-D Scatterer Models

A 3D scatterer model from a sequence of ISAR images can be formed in three steps. In the initialization step, the images are aligned in range and the scatterer Doppler measurements are extracted. Scatterers are associated between frames based on their intensity. In the second step, the 3D scatterer positions and ship motion are estimated based on the Doppler measurements. Because there is not a unique solution to the underlying equation for this process, linear transformations may be required to map the scatterers into the correct coordinate system. In the final step of relabeling, the algorithm iterates on a solution to compensate for poor labeling in the initialization step. The weakest scatterers usually have poor position estimates, so we may want to remove them from the 3-D model before classification. Classification using these 3-D models can be done in four steps. First, a 3D model can be estimated from the images of interest. Then, the new model must be aligned into the same coordinate system as the library model. Because it is difficult to compare the models point by point, the library model should be smoothed by replacing discrete points by a mixture of Gaussian clusters. Bidirectional match scores are calculated between the two models and the unknown ship is classified into the class giving the highest similarity score [70].

14.5.3 Candidate Features

Challenges inherent in defining features from inverse synthetic aperture radar images include motion compensation while imaging multiple targets and aspect dependence on the position of the centroid, rotation of the object, and placement of scatterers. The following sections describe a few feature extraction algorithms and typical feature sets that have been developed to deal with these challenges.

14.5.3.1 Edge Detection and Watershed Segmentation

Shape extraction can be performed using several common image processing steps. First, for a high SNR ISAR image, a simple mean threshold can be applied to segment the target from noise and clutter. Shape extraction can then be performed by marking the targets defined by segmentation and applying the watershed algorithm. In this algorithm, the image is considered to be a topographic relief with grayscale values as the altitude; individual features in the image are defined by where water would pool based on the terrain. To segment a target in an ISAR image, the watershed transformation is applied to the gradient image [53]. Each pixel in the catchment basins may then be replaced by the minimum image gradient contained within, and the watershed process can be iterated upon to merge catchment basins until the desired object is obtained [53]. Shape descriptors resulting from this segmentation should be accurate, compact, and invariant to translation, rotation, and scaling. Descriptors that fit these criteria include Fourier descriptors, where the Fourier transform is applied to the normalized shape boundary (normalized by the number of points in the boundary) [53].

14.5.3.2 Hough Transform

Edge detection can be performed on the absolute values of the range profile histories (in cross-range and down-range) of multiple targets using Sobel, Roberts, Canny, or Prewitt methods [73]. Using the edge information from range profile histories, the Hough transform will estimate the slope of the range profile history of each target [73]. The Hough transform finds the angle and intercept of a line, and then multiple targets can be separated based on the angles with the maximum Hough transform values [73]. The now segregated range profile histories can be aligned by minimizing the 1-D entropy function between the two range profiles or, in a less time-consuming process, by using the slopes to collapse all the scattering centers on the same line to the position of the first one [73].

14.5.3.3 Wavelet and Adaptive Contourlet Transform

Wavelet transform coefficients can be used as image features for classification, but are not good at capturing the geometry of image edges. Contourlets are a form of geometric wavelet that can better find the features inherent in the contours, edges, textures, and background of a SAR image; however, they may not capture high frequency details of the image. By combining the properties of adaptive contourlets and wavelet-based contourlets, a truncated contourlet decomposition tree can be constructed as a directional and multiscale representation (or decomposition representing intrinsic geometrical structures) of an image using a genetic algorithm for optimization. Some statistical features of the tree nodes can serve as features for classification, including energy, kurtosis, skewness, median, and mean [74].

14.5.3.4 Log-Spiral Transform

The log-spiral transform can be used to rotate the object and translate the centroid to a characteristic mode, which is aspect independent [75]. The outline of the target in the 2-D Euler plane is mapped to a calculation plane by the log-spiral transform, such that $u = \ln((x - x_c)^2 + (y - y_c)^2)/2$ and $v = \tan^{-1}((y - y_c)/(x - x_c))$ where (x_c, y_c) is the centroid of the target. By way of this transformation, the rotation of the target around the centroid is changed to translation of a curve along the horizontal and the range between a point on the target outline and the centroid becomes variation of the curve along the

ordinate, u [75]. Thus, the log-spiral transform is translation, rotation, and scale invariant when the centroid is moved to a baseline point for each image [75].

14.5.3.5 Ship Features

Features for land targets have been previously discussed in the SAR section. However, images of ships and other watercraft at sea are often linked specifically to ISAR imagery; therefore, useful features for ship classification are discussed.

Ships at sea are a common target for ISAR imaging with multiple aspect angles and image orientations captured due to the changing direction of the angular velocity vector as the ship moves along the waves. Some of the most useful classification features for large ships are ship length, number of masts, and mast location [67]. For small ships, these features can be supplemented with width or height, shape, guns, and gun location [67]. Scaled and oriented wire grids from candidate ship libraries can be compared to ISAR ship images to aid in classification [67, 69]. To determine the orientation of a ship in an ISAR image, features such as the centerline (using maximum peak of Hough transform of partially segmented image), locations of the bow and stern, how much plan information is present (peaks in the Hough transform at the angle parallel to centerline), how much profile information is present, the width of the plan at either end, and the width of the outline at either end can be used [69]. In a process sometimes referred to as mensuration, the locations of major structures, such as masts, superstructure breaks, cranes, and weapons, are found along the bow-stern axis and normalized relative to the range extent of the ship; these locations can be determined by looking at the number of target pixels extending in the Doppler direction beyond either side of the deck edges for each range cell [69]. Other useful features in ISAR ship imagery include: straight, curved, or rounded stern; pole, lattice, or solid mast; target range extent; percentage of the target range extent covered by superstructure; number of major uprights; the ratio of the highest mast to the median height of the profile; and the correlation of the five best profiles with a stored characteristic profile shape [69]. Because many of these features require maximum height profile information, ship ISAR images are selected for evaluation based on a histogram of the number of target pixels per range cell on the side of the target with the most pixels outside of the plan boundary. Vertical streaks caused by transient or rapid motions internal to the target (e.g., moving machinery) must be removed before extracting relevant features from the image. This can be done by counting the number of pixels above a threshold in cross-range for each range cell, by flagging range cells with more than 66% of their pixels above this threshold, by finding pixels in these flagged ranges that are bright while their neighbors are not, and by reducing the intensity of such pixels by 30% [69].

14.6 | PASSIVE RADAR ATR

While traditional monostatic radar systems like those commonly used to perform SAR and ISAR include both transmitters and receivers, passive radar systems contain only receivers. They rely on “illuminators of opportunity” already present in the environment (e.g., antennas used to transmit signals for television, cell phones, or FM radio) to illuminate potential targets [111–114].

The potential benefits of this technology are numerous. Since passive radar systems never emit energy (relying only on signals already present), they can operate covertly; targets have no way of knowing that they are being detected, tracked, and classified by passive

radar systems. Furthermore, since the majority of the cost associated with traditional radar systems is associated with the transmitter, passive radar systems are comparatively inexpensive to build and maintain. Although passive radar is not inherently restricted to a particular frequency band, the allocation of bandwidth for common illuminators of opportunity (e.g., commercial TV and FM radio) serendipitously relegates most passive radar systems to low-frequency regimes. Consequentially, their performance is not prone to substantial degradation in inclement weather, providing a tactical advantage. Another fortunate consequence of this association with low-frequency sources is that the longer wavelengths are well-suited for many ATR applications, such as aircraft classification [115–116]. Recall that at X-band (e.g., 10 GHz), a typical fighter plane is roughly 500 wavelengths long, while at FM-band frequencies (e.g., 100 MHz), the same fighter is only five wavelengths long. Where the X-band radar cross section of the aircraft is extremely sensitive to small perturbations in aircraft orientation, the FM-band radar cross section of the aircraft varies slowly and smoothly with changes in aspect, providing a feature that varies enough to allow classification but not so much as to require high precision prediction of aircraft orientation, which can be difficult to achieve.

Even so, a number of unique challenges are associated with passive radar. The following sub-sections describe these challenges in the context of the unified framework for ATR.

14.6.1 Step 1: Identify the Target Set

First and foremost, the target sets that are candidates for passive radar ATR are largely dictated by the illuminators of opportunity in the deployment area. As a result, these systems are most commonly associated with urban areas in developed countries that have stringent communications regulations. Deployment in under-developed countries is difficult due to the high noise figures that result from lax emissions regulations. Deployment in rural (but regulated) areas is challenging due to the lack of available illuminators of opportunity. Hence, when identifying the target set, the passive radar ATR designer must begin by considering the available sources in the deployment area. These, in turn, dictate the operational region (i.e., the region in which the signals from multiple transmitters can be reliably received). The target sets are thus restricted to targets that will pass through this region and generate sufficient signal-to-noise ratio to be detected. The physical size of the targets (relative to the size of the illuminators' wavelengths) must also be considered.

14.6.2 Step 2: Select the Feature Set

The feature sets available to passive radar ATR systems are also somewhat limited. Radar cross section and target kinematics can be estimated over time by the passive radar ATR system, but additional features are difficult to collect since the system is unable to control transmitted waveforms.

14.6.3 Step 3: Observe the Feature Set

Since direct path signals (i.e., those moving directly from the transmitters to the receiver without hitting the target) are orders of magnitude stronger than the desired target signals, deployment of passive radar receivers to adequately observe the feature set requires careful planning. The transmitters' power and locations are typically known and the associated spikes can be removed from the ambiguity function via signal processing, but the direct path signals raise the noise floor, potentially swamping the indirect target signals. Hence, passive

radar receivers typically use sophisticated beamforming to place nulls in the direction of direct path signals while optimizing gain towards the operational regions. This allows them to adequately detect and observe targets.

The nature of passive radar measurements also complicates the observation and prediction of kinematic target features. While traditional radar systems excel at measuring the range to the target, passive radar systems are best at measuring direction-of-arrival (DOA) and Doppler. Hence, translation into accurate state estimates in Cartesian coordinate systems (often required to collect kinematic features) is a challenge. In practice, this is addressed through a combination of employing multiple receivers [117] and using sophisticated signal processing and tracking algorithms [111].

14.6.4 Step 4: Test the Feature Set

Once features have been observed, likelihood-based schemes [100–105] to compare them to a target library are common. For example, much of the literature in this area focuses on use of radar cross sections for classification of aircraft. In this case, the power at the passive radar receiver is modeled as

$$P_{RECEIVER} = \left(\sqrt{P_{TARGET}} + w_R \right)^2 + w_I^2 \quad (14.16)$$

where P_{TARGET} is the real component of the received power due to the target and w is zero-mean additive white Gaussian noise with real, w_R , and imaginary, w_I , parts [100–105]. This leads to a Rician likelihood model, whose probability density function is given by

$$p_x(x) = \frac{x}{\sigma_w^2} \exp \left[- \left(\frac{x^2 + s^2}{2\sigma_w^2} \right) \right] I_0 \left(\frac{xs}{\sigma_w^2} \right) \quad (14.17)$$

where x is the observed voltage at the receiver, s is the predicted voltage at the receiver for the target from the library, σ_w^2 is the noise power, and I_0 is the modified Bessel function of the first kind [102]. If each of N_{MEAS} measurements is assumed to be independent, the loglikelihood becomes

$$\ln(p_x(\bar{x})) = \sum_{i=1}^{N_{MEAS}} \ln \left(\frac{x_i}{\sigma_w^2} \right) + \ln \left[I_0 \left(\frac{x_i s_i}{\sigma_w^2} \right) \right] - \left(\frac{x_i^2 + s_i^2}{2\sigma_w^2} \right) \quad (14.18)$$

Since the first term in the loglikelihood equation does not vary with the elements in the target library, the score, S , used to identify the target can be reduced to

$$S(\bar{x}) = \sum_{i=1}^{N_{MEAS}} \ln \left[I_0 \left(\frac{x_i s_i}{\sigma_w^2} \right) \right] - \left(\frac{x_i^2 + s_i^2}{2\sigma_w^2} \right) \quad (14.19)$$

The target from the library resulting in the maximum score wins.

14.7 | HIGH-RESOLUTION RANGE PROFILES

High-resolution range profiles (HRRPs) are based on returns from high-resolution radars (HRRs) or ultra-high-resolution radars. In this paradigm, the HRRP is the time domain response to interrogating the target with a set of high-range resolution pulses [76]. The pulsewidths used to interrogate the target are selected to produce many returns, which divide targets of interest into multiple returns, provide information about the target scatterers, and by extension, the target's composition.

When first introduced, HRRPs seemed to be a promising new approach and spawned numerous publications. However, HRRPs compress target responses from three-dimensions into a single dimension (range), making them fundamentally non-unique. Since the complex exponential signals (constituting the sole available feature in this scheme) can combine constructively or destructively during the dimension reduction, different targets could theoretically provide identical HRRPs. As a result, HRRPs are most commonly used today as secondary features in the context of more comprehensive ATR schemes. For example, HRRPs are commonly exploited in SAR and moving target indication (MTI) applications when targets cross [77–79], as they can help resolve the targets more quickly than is possible in these other venues. For this reason, we skip the first step in the unified framework (identifying the target set) and move onto Step 2, feature selection.

14.7.2 SET 2: SELECT THE FEATURE Set

The HRRPs produced from HRRs are the main feature for this approach to ATR. Tait [80] defines high-resolution range profiles as a “representation of the time domain response of the target to a high-range resolution radar pulse.” Another way to think of it is that HRRPs are a 1-D signature, and are the result of collapsing 3-D data.

In understanding how HRRPs are formed, the crux of why these are not as commonly used for ATR becomes evident. For an example we consider standing on the ground and looking at an airplane. Depending on your viewing angle, a plane (and its scatters) can appear to have different strengths and be at different relative locations. The problem becomes exacerbated when you consider that the radar cannot assume it is in the same plane as its target of interest, but must take into account all 360° at a variety of different heights both above and below the target. Further, different models and types of airplanes all have different scatters at different aspect angles.

When designing an ATR system to work on HRRPs, one of the first decisions is what data should be modeled and stored? It is not clear that there is a right or wrong answer, but instead should be based on your problem, including the appearance of your targets and scene and your computing resources. Jacobs and O’Sullivan [76, 81] discuss how to model range profiles, presenting two options. The traditional approach treats range profiles as deterministic, as in the classic electromagnetic sense. Their second option treats range profiles as stochastic.

14.7.2.1 Saving the Feature Set

Coupled with the decision of how to model the HRRP is how to create your range profile database. The first decision is whether to store 2-D or 3-D data [80] in your range profile database. A generic 2-D table has target type versus aspect angle. The values in the table are normally averages of range profiles or a set of range profile features. Extending to 3-D, the third dimension is normally elevation angles of the target. Along with this decision, is how to store a range profile. One option is to store the entire feature vector, or whole range profile [82], while another option is to store only the amplitude and location of the signature peaks [83]. Libby and Maybeck [84] store HRRPs as a range-bin length vector for each aspect angle, polarization, center frequency and bandwidth of interest.

14.7.3 Step 3: Observe the Feature Set

The observation of the range profiles is based on the waveforms that are to be used. Tait includes a detailed discussion on the waveform decision in his chapter on HRRPs [80].

Some waveform decisions include whether or not to use stepped frequencies, pulse compression, stretch processing and bandwidth segmentation. Additionally, Tait presents a good summary of the effect of waveform choice on range window, time on target, range ambiguities and operation in clutter environments.

An additional consideration for feature observation is that performance for HRRPs is largely based on knowing the aspect angle of the target of interest, since scatters can change significantly. To address this problem, normally a target is traced for a short time to determine its direction before doing range-profile measurements. With an approximate direction, the number of aspect angles to be considered is reduced.

An equally challenging issue is clutter suppression, which is often necessary to distinguish the target. Kahler and Blasch [85] discuss several techniques for mitigating the effects of clutter, including Doppler filtering, space-time adaptive processing (STAP) and displaced phase center arrays (DPCA). The Doppler filtering approach can be used with adaptive radars that are capable of detecting the distribution of clutter and adjusting the radar appropriately to minimize the signal-to-clutter ratio. STAP uses Doppler frequency, platform velocity and direction of arrival information to reduce clutter. Displaced phase center arrays compensate for radar motion to reduce the Doppler spread of ground clutter.

Additional difficulties include the existences of multiple or extended targets, which can go against the initial assumptions of the ATR system. Some HRR approaches are more susceptible to problems related to moving targets, and thus require the data to be motion compensated, or mo-comped (a process of measuring target velocity and adjusting the phase accordingly). Additionally, data compression must occur for real systems [86]. Unfortunately, compression is expensive, and involves calibration and measurement errors. Unpredictable experimental parameters and a lack of realistic target data for all configurations and scenarios also add to the complexity of the data compression problem.

14.7.4 Step 4: Test the Feature Set

A number of techniques are given in the literature for testing HRRPs, even if only as a secondary feature in a broader ATR scheme. Mitchell and Westerkamp [82] explain the traditional approach as a constrained quadratic classifier. To implement this approach, compute an ensemble mean and variance estimate for each range bin in the HRRP signatures. Then discriminate based on the observation and statistics (mean and variance). This can also be implemented with a mean-squared error to eliminate the variance terms. They also present a statistical feature-based classifier [82–83] method, which builds on an idea originally developed for air target identification and is implemented by looking at probabilities calculated with a peak location probability function. Similarly, Jacobs and O’Sullivan [81, 76] use a likelihood approach, combining dynamics based on target orientation and likelihoods of pre-determined profiles, given the orientation. Still others [84–89] choose a Mahalanobis distance metric that treats HRRR measurements as a vector of Gaussian elements.

Layne [90] presents a multiple model estimator (MME) approach with multiple Kalman Filters (KF). His work builds on previous work that considered MMEs for SAR measurements [76, 91–92]. Layne’s approach is to use one KF for each target type and estimate kinematics (including position, velocity, and possibly acceleration). He then is able to predict the target based on likelihoods from the various filters. The filters are designed to

take a target cross sectional length measurement from the HRRP as an input and map it to a set of aspect angles. Those angles are then compared with the kinematics measurements to find the angle that is most consistent, and thus minimizes the filter residual.

Several authors have taken a more mathematical approach and started with the eigenvalue or singular value decomposition (SVD) of the HRRPs, followed by eigenvalue or singular value template matching [85, 93–97]. These matrix decomposition approaches exploit structure in the data. In general, a dominant eigenvector accounts for 90% of the target's energy, and the range and angle bias spaces are numerically decoupled in the left and right eigenvectors [98]. Further, using Principal Component Analysis (PCA) to eliminate eigenvectors of small eigenvalues can lead to clutter suppression [93]. Likewise the SVD allows for projection of the HRRP onto orthogonal basis spaces (this is Karhunen-Loeve transformation or again, PCA), like in the Baseline Automated Recognition of Targets (BART) proposed by Kahler and Blasch [96].

Nguyen, et al. [99] proposed High Definition Vector Imaging (HDVI), a super-resolution technique. They apply HDVI to HRRPs before using the profiles for ATR classification. They argue this approach requires fewer computations than weighted fast Fourier transforms and can improve recognition over traditional image processing techniques.

14.8 | SUMMARY

A wide variety of radar-based ATR systems and applications have been developed and documented in open literature. As different as these may seem on the surface, they all tend to revolve around four basic steps, labeled in this chapter as the unified framework for ATR. First, the target set must be identified. Then the feature set must be selected, observed, and tested to identify the targets. The most challenging parts of the problem tend to be predicting all likely target variations and discerning a feature set that sufficiently segregates the target set while being readily observable.

14.9 | FURTHER READING

Recommended further reading focus on the computational aspects of database formation and access, which are outside the scope of this chapter. A fundamental tradeoff exists between the desired performance (and associated degree of fidelity required in the target database) and the amount of available computing power. More information can improve robustness, but it also implies increased computational demands. One approach is to employ a traditional clustering technique (such as k-means clustering algorithm [86]) to reduce the number of signatures in the database. High performance computing may also be required to deliver results that are both sufficiently reliable and timely.

A significant number of papers address the problems of ATR in a theoretical sense, but actual implementation of their ideas is not always straightforward. For example, [77] [78] assume their MTI tracker can compute target aspect angles within $\pm 15^\circ$. This assumption means that fewer angles need to be stored in their database. They also assume that their templates and observations can be aligned and normalized. Other authors make assumptions about the data collection, including that the flight path of the radar is constant velocity and roll angle of 0° .

14.10 | REFERENCES

- [1] L. M. Novak, S. D. Halversen, G. J. Owirka and M. Hiatt, "Effects of Polarization and Resolution on the Performance of a SAR Automatic Target Recognition System," *The Lincoln Laboratory Journal*, vol. 8, no. 1, pp. 49–68, 1995.
- [2] M. Vespe, C. J. Baker and H. D. Griffiths, "Automatic target recognition using multi-diversity radar," *IET Radar Sonar Navig.*, vol. 1, no. 6, pp. 470–478, 2007.
- [3] A. E. Catlin, "System Comparison Procedures for Automatic Target Recognition Systems," Air Force Institute of Technology, Wright Patterson AFB, OH, 1997.
- [4] R. Hummel, "Model-Based ATR Using Synthetic Aperture Radar," in *IEEE International Radar Conference*, 2000.
- [5] T. D. Ross and J. C. Mossing, "The MSTAR Evaluation Methodology," in *Algorithms for Synthetic Aperture Radar Imagery VI Proceedings of the SPIE*, Orlando, FL, 1999.
- [6] T. D. Ross, L. A. Westerkamp, D. A. Gadd and R. B. Kotz, "Feature and Extractor Evaluation Concepts for Automatic Target Recognition (ATR)," Wright Laboratory, Wright-Patterson AFB, OH, 1995.
- [7] L. M. Novak, G. J. Owirka and A. L. Weaver, "Automatic Target Recognition Using Enhanced Resolutions SAR Data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 1, pp. 157–175, 1999.
- [8] L. Westerkamp, S. Morrison, T. Wild and J. Mossing, "High-resolution synthetic aperture radar experiments for ATR development and performance prediction," in *Proceedings of SPIE: Algorithms for Synthetic Aperture Radar Imagery IX*, 2002.
- [9] A. K. Mishra and B. Mulgrew, "Multipolar SAR ATR: Experiments with the GTRI Dataset," in *IEEE Radar Conference*, 2008.
- [10] European Aeronautic Defence and Space Company, "Generation of Synthetic SAR Imagery for ATR Development," European Aeronautic Defence and Space Company, Munich, Germany, 2005.
- [11] C. J. Beaudoin, A. J. Gatesman, R. H. Giles, J. Waldman and W. E. Nixon, "A 3D Polar Processing Algorithm for Scale Model UHF ISAR Imaging," University of Massachusetts Lowell, Submillimeter-Wave Technology Laboratory, Lowell, MA, 2006.
- [12] L. M. Novak, G. J. Owirka and C. M. Netishen, "Performance of a High-Resolution Polarimetric SAR Automatic Target Recognition System," vol. 6, no. 1, pp. 11–24, 1993.
- [13] A. K. Mishra, "Validation of PCA and LDA for SAR ATR," in *IEEE Region 10 Conference*, 2008.
- [14] A. K. Mishra and B. Mulgrew, "Multipolar SAR ATR: Experiments with the GTRI Dataset," in *IEEE Radar Conference*, 2008.
- [15] A. K. Mishra and B. Mulgrew, "Bistatic SAR ATR," *IET Radar Sonar Navig.*, vol. 1, no. 6, pp. 459–469, 2007.
- [16] E. Seidenberg and H. Schimpf, "Aspects of Automatic Target Recognition with a Two-Frequency Millimeter Wave SAR," in *Radar Sensor Technology V Proceedings of SPIE*, 2000.
- [17] V. Sabio, "Target Recognition in Ultra-Wideband SAR Imagery," U.S. Army Research Laboratory, Adelphi, MD, 1994.
- [18] A. J. Bennett and A. Currie, "The Use of High Resolution Polarimetric SAR for Automatic Target Recognition," in *Proceedings of SPIE*, 2002.

- [19] Q. H. Pham, A. Ezekiel, M. T. Campbell and M. J. Smith, "A New End-to-End SAR ATR System," in *Algorithms for Synthetic Aperture Radar Imagery VI Proceedings of the SPIE*, Orlando, FL, 1999.
- [20] R. Chellappa, "Advanced Automatic Target Recognition," Center for Automation Research, University of Maryland, College Park, MD, 1998.
- [21] S. K. Wong, "High Range Resolution Profiles as Motion-Invariant Features for Moving Ground Targets Identification in SAR-Based Automatic Target Recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 1017–1039, 2009.
- [22] L. I. Voicu, R. Patton and H. R. Myler, "Multi-criterion vehicle pose estimation for SAR-ATR," in *Algorithms for Synthetic Aperture Radar Imagery VI Proceedings of the SPIE*, Orlando, FL, 1999.
- [23] J. A. O'Sullivan, M. D. DeVore, V. Kedia and M. I. Miller, "SAR ATR Performance Using a Conditionally Gaussian Model," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 1, pp. 91–108, 2001.
- [24] T. Kempf, M. Peichl, S. Dill and H. Suess, "Highly resolved turntable ISAR signature extraction for ATR," in *Proceedings of SPIE*, 2009.
- [25] K. Xue and S. Sink, "Synthetic Aperture Radar (SAR) Automatic Target Recognition (ATR) Parametric Study," Wright State University, Department of Electrical Engineering, Dayton, OH, 2003.
- [26] J. C. Mossing, T. D. Ross and J. Bradley, "An Evaluation of SAR ATR Algorithm Performance Sensitivity to MSTAR Extended Operating Conditions," Air Force Research Laboratory, Wright-Patterson AFB, OH, 1998.
- [27] L. M. Novak, "State-of-the-Art of SAR Automatic Target Recognition," in *IEEE International Radar Conference*, 2000.
- [28] T. Kempf, M. Peichl, S. Dill and H. SuB, "Influence of Camouflage on the ATR Performance for Highly Resolved ISAR-Images of Relocatable Targets," DLR (German Aerospace Center), Microwave and Radar Institute, Wessling, Germany, 2005.
- [29] G. R. Benitz, "High-Definition Vector Imaging," *Lincoln Laboratory Journal*, vol. 10, no. 2, pp. 147–160, 1997.
- [30] M. Soumekh, "Adaptive Processing of SAR Data for ATR," SUNY-Buffalo, Amherst, NY, 2005.
- [31] S. Worrell and M. Soumekh, "Automatic Target Recognition in SAR Using Digitally-Spotlighted Phase History Data," in *Proceedings of SPIE: Algorithms of Synthetic Aperture Radar Imagery VII*, 2000.
- [32] A. Rosenfeld, "Advanced Automatic Target Recognition," University of Maryland, College Park, MD, 1997.
- [33] Q. H. Pham, A. Ezekiel, M. T. Campbell and M. J. Smith, "A New End-to-End SAR ATR System," in *Proceedings of the SPIE: Algorithms for Synthetic Aperture Radar Imagery VI*, Orlando, FL, 1999.
- [34] K. C. Chang, "Detection and Classification of Synthetic Aperture Radar Targets," Center of Excellence in C3I, George Mason University, Fairfax, VA, 1997.
- [35] Q. Li, J. Wang, B. Zhao, F. Luo and X. Xu, "An Automatic Target Recognition System based on SAR Image," in *MIPPR 2009: Automatic Target Recognition and Image Analysis Proceedings of SPIE*, 2009.
- [36] N. M. Sandirasegaram, "Automatic Target Recognition in SAR Imagery using a MLP Neural Network," Ottawa, Ontario, 2002.

- [37] N. M. Sandirasegaram, "Spot SAR ATR Using Wavelet Features and Neural Network Classifier," Ottawa, Ontario, 2005.
- [38] Y. Wang, P. Han and R. Wu, "Analysis of PC Number Selection in SAR ATR," in *1st Asian and Pacific Conference on Synthetic Aperture Radar*, 2007.
- [39] X. Lu, P. Han and R. Wu, "Two-Dimensional PCA for SAR Automatic Target Recognition," in *1st Asian and Pacific Conference on Synthetic Aperture Radar*, 2007.
- [40] L. Hu, J. Liu, H. Liu, B. Chen and S. Wu, "Automatic Target Recognition based on SAR images and Two-Stage 2DPCA features," in *1st Asian and Pacific Conference on Synthetic Aperture Radar*, 2007.
- [41] L. Maokuan, G. Jian, D. Hui and G. Xin, "SAR ATR based on Support Vector Machines and Independent Component Analysis," in *International Conference on Radar*, 2006.
- [42] N. S. Subotic, J. D. Gorman and S. Welby, "Hybrid Template- and Model-Based ATR Formulation," US Army Research Laboratory, Adelphi, MD, 1999.
- [43] O. Horsfield and D. Blacknell, "Utilising feedback in adaptive SAR ATR systems," in *Algorithms for Synthetic Aperture Radar Imagery XVI Proceedings of SPIE*, 2009.
- [44] H. Seidel, C. Stahl, P. Knappe and P. Hurst, "Generation of Synthetic SAR Imagery for ATR Development," Munich, Germany, 2005.
- [45] T. O. Binford, B.-H. Wang and T. S. Levitt, "Context and Quasi-Invariants in Automatic Target Recognition (ATR) with Synthetic Aperture Radar (SAR) Imagery," Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, 2002.
- [46] M. Cetin, W. C. Karl and D. A. Castanon, "Analysis of the Impact of Feature-Enhanced SAR Imaging on ATR Performance," in *Proceedings of the SPIE*, 2002.
- [47] L. C. Potter and R. L. Moses, "Attributed Scattering Centers for SAR ATR," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 79–91, 1997.
- [48] D. Waagen, J. Pecina and R. Pickens, "Evolving Spatially-Localized Projection Filters for SAR Automatic Target Recognition," in *Proceedings of the 7th International Conference on Evolutionary Programming VII*, 1998.
- [49] Y. Chen, E. Blasch, H. Chen, T. Qian and G. Chen, "Experimental Feature-Based SAR ATR Performance Evaluation under Different Operational Conditions," in *Proceedings of the SPIE*, 2008.
- [50] D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge, "Comparing Images Using the Hausdorff Distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.
- [51] H. Yuankui and Y. Yiming, "Automatic Target Recognition of ISAR Images Based on Hausdorff Distance," in *1st Asian and Pacific Conference on Synthetic Aperture Radar*, 2007.
- [52] J. A. Saghri, "SAR Automatic Target Recognition Using Maximum Likelihood Template-based Classifiers," in *Applications of Digital Image Processing XXXI Proceedings of the SPIE*, 2008.
- [53] A. Toumi, B. Hoeltzener and A. Khenchaf, "Using Watersheds segmentation on ISAR image for automatic target recognition," in *2nd International Conference on Digital Information Management*, 2007.
- [54] K. Ikeuchi, R. Collins, T. Shakunaga and K. Ohba, "Adaptive Model Based ATR System," Carnegie Mellon University, Pittsburgh, PA, 1996.
- [55] Q. Zhao, D. Xu and J. C. Principe, "Pose Estimation for SAR Automatic Target Recognition," in *Proceedings of Image Understanding Workshop*, 1998.

- [56] M. Martorella, E. Giusti, A. Capria, F. Berizzi and B. Bates, "Automatic Target Recognition by Means of Polarimetric ISAR Images and Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3786–3794, 2009.
- [57] Y. de Villers, "Development and Test of a Millimetre-Wave ISAR Images ATR System," Defence R D Canada, Valcartier Val-Belair, Canada, 2005.
- [58] H.-C. Chiang, R. L. Moses and L. C. Potter, "Classification Performance Prediction Using Parametric Scattering Feature Models," in *Algorithms for Synthetic Aperture Radar Imagery VII Proceedings of SPIE*, 2000.
- [59] C. Nilubol, Q. H. Pham, R. M. Mersereau, M. J. Smith and M. A. Clements, "Hidden Markov Modelling for SAR Automatic Target Recognition," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA, 1998.
- [60] S. Papson and R. M. Narayanan, "Applying Target Shadow Models for SAR ATR," in *Signal Processing, Sensor Fusion, and Target Recognition XVI Proceedings of the SPIE*, 2007.
- [61] W. Xiong and L. Cao, "Automatic Target Recognition Based on Rough Set-Support Vector Machine in SAR Images," in *2009 International Joint Conference on Computational Sciences and Optimization*, 2009.
- [62] Y. Wang, P. Han, X. Lu, R. Wu and J. Huang, "The Performance Comparison of Adaboost and SVM Applied to SAR ATR," in *International Conference on Radar*, 2006.
- [63] W. Middelmann, A. Ebert and U. Thoennessen, "Automatic Target Recognition in SAR Images Based on a SVM Classification Scheme," in *ICANNGA*, 2007.
- [64] A. Mahalanobis, D. W. Carlson and B. Vijaya Kumar, "Evaluation of MACH and DCCF correlation Filters for SAR ATR using MSTAR public data base," in *Algorithms for Synthetic Aperture Radar Imagery V Proceedings of SPIE*, Orlando, FL, 1998.
- [65] M. Alkanhal, B. Vijaya Kumar and A. Mahalanobis, "Improved clutter rejection in automatic target recognition (ATR) synthetic aperture radar (SAR) imagery using the extended maximum average correlation height (EMACH) filter," in *Proceedings of SPIE*, 2000.
- [66] D. W. Carlson, B. V. Vijaya Kumar, R. R. Mitchell and M. Hoffelder, "Optimal trade-off distance classifier correlation filters (OTDCCFs) for synthetic aperture radar automatic target recognition (SAR ATR)," in *Proceedings of the SPIE*, 1997.
- [67] J. C. Kirk Jr., "Options for Time-Frequency Processing in ISAR ATR," in *Proceedings of SPIE*, 2001.
- [68] K. A. Melendez and J. R. Bennett, "ISAR Target Parameter Estimation with Application for Automatic Target Recognition," in *SPIE Conference on Radar Processing, Technology, and Applications*, 1998.
- [69] S. Musman, D. Kerr and C. Bachmann, "Automatic Recognition of ISAR Ship Images," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1392–1404, 1996.
- [70] T. Cooke, M. Martorella, B. Haywood and D. Gibbins, "Use of 3D Ship Scatterer Models from ISAR Image Sequences for Target Recognition," *Digital Signal Processing*, vol. 16, no. 5, pp. 523–532, 2006.
- [71] Martorella, M.; Giusti, E.; Berizzi, F.; Dalle Mese, E., "Automatic Target Recognition of Terrestrial Vehicles based on Polarimetric ISAR Image and Model Matching," in *IEEE Radar Conference*, 2008.
- [72] R. G. Baird and W. E. Nixon, "Development and assessment of a complete ATR algorithm based on ISAR Euler imagery," in *Proceedings of SPIE*, 2007.

- [73] S. H. Park, K. K. Park, J. H. Jung and H. T. Kim, "ISAR Imaging of Multiple Targets Using Edge Detection and Hough Transform," *J. of Electromagn. Waves and Appl.*, vol. 22, pp. 365–373, 2008.
- [74] S. Yang, M. Wang and L. Jiao, "Radar target recognition using contourlet packet transform and neural network approach," *Signal Processing*, vol. 89, pp. 394–409, 2009.
- [75] W. Ning, W. Chen and X. Zhang, "Automatic Target Recognition of ISAR Object Images Based on Neural Network," in *IEEE Int. Conf. Neural Networks & Signal Processing*, 2003.
- [76] S. P. Jacobs and J. A. O'Sullivan, "High resolution radar models for joint tracking and recognition," in *IEEE National Radar Conference*, 1997.
- [77] R. Williams, J. Westerkamp, D. Gross and A. Palomino, "Automatic target recognition of time critical moving targets using 1D high range resolution (HRR) radar," in *The Record of the 1999 IEEE Radar Conference*, 1999.
- [78] R. Williams, J. Westerkamp, D. Gross and A. Palomino, "Automatic target recognition of time critical moving targets using 1D high range resolution (HRR) radar," *IEEE Aerospace and Electronic Systems Magazine*, vol. 15, no. 4, pp. 37–43, 2000.
- [79] M. T. Fennell and R. P. Wishner, "Battlefield awareness via synergistic SAR and MTI exploitation," *IEEE Aerospace and Electronic Systems Magazine*, vol. 13, no. 2, pp. 39–43, 1998.
- [80] P. Tait, "Introduction to Radar Target Recognition," *IEE Radar, Sonar and Navigation series*, vol. 18, 2005.
- [81] S. P. Jacobs and J. A. O'Sullivan, "Automatic target recognition using sequences of high resolution radar range-profiles," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 364–381, 2000.
- [82] R. A. Mitchell and J. J. Westerkamp, "Robust statistical feature based aircraft identification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 3, pp. 1077–1094, 1999.
- [83] R. A. Mitchell and J. J. Westerkamp, "Statistical feature based target recognition," in *Proceedings of the IEEE 1998 National Aerospace and Electronic Systems Conference*, 1998.
- [84] E. W. Libby and P. S. Maybeck, "Sequence comparison techniques for multisensor data fusion and target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 1, pp. 52–65, 1996.
- [85] B. Kahler and E. Blasch, "Impact of HRR radar processing on moving target identification performance," in *12th International Conference on Information Fusion*, 2009.
- [86] B. Ulug, S. C. Ahalt and R. A. Mitchell, "Efficient ATR using compression," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1199–1211, 1997.
- [87] C. V. Stewart, V. J. Larson and J. D. Halsey, "Comparison of classification approaches for high-range resolution radar," in *Proceedings of SPIE*, 1992.
- [88] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Boston, MA: Academic Press, 1990.
- [89] H. Mieras and T. Cichocki, "Airframe identification results using high range resolution," in *Joint Service Combat Identification System Conference*, 1990.
- [90] J. R. Layne, "Automatic Target Recognition and Tracking Filter," in *SPIE Aerosense - Small Targets*, 1998.
- [91] E. Libby and P. S. Maybeck, "Application of sequence comparisons methods to multisensor data fusion and target recognition," *IEEE Aerospace and Electronic Systems Magazine*, pp. 52–65, January 1996.

- [92] E. Libby, "Application of sequence comparison methods to multisensor data fusion and target recognition," Air Force Institute of Technology, 1993.
- [93] A. S. Paul and A. K. Shaw, "Robust HRR radar target identification by hybridization of HMM and eigen-template-based matched filtering," in *Proceedings of SPIE*, 2003.
- [94] A. K. Shaw, "Automatic target recognition using high-range resolution data," Wright State University, Dayton, 1998.
- [95] K. Shaw and V. Bhatnagar, "Automatic target recognition using eigen-templates," in *Proceedings of SPIE*, 1998.
- [96] B. Kahler and E. Blasch, "Robust multi-look HRR ATR investigation through decision-level fusion evaluation," in *11th International Conference on Information Fusion*, 2008.
- [97] B. V. Kumar and M. Alkanhal, "Eigen-extended maximum average correlation height (EEMACH) filters for automatic target recognition," in *Proceedings of SPIE*, 2001.
- [98] A. K. Shaw and R. Vashist, "Automatic Target Recognition using Eigen Templates," in *Proceedings of SPIE Synthetic Aperture Radar Imagery V*, Orlando, FL, 2000.
- [99] D. Nguyen, G. Benitz, J. Kay, B. Orchard and R. Whiting, "Superresolution HRR ATR with high definition vector imaging," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 4, pp. 1267–1286, 2001.
- [100] S. Herman, *A Particle Filtering Approach to Joint Passive Radar Tracking and Target Classification*, Urbana, IL: Doctoral Dissertation, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 2002.
- [101] S. Herman and P. Moulin, "A Particle Filtering Approach to Joint Radar Tracking and Automatic Target Recognition," in *Proceedings of the IEEE Aerospace Conference*, Big Sky, MT, 2002.
- [102] L. Ehrman, *An Algorithm for Automatic Target Recognition Using Passive Radar and an EKF for Estimating Aircraft Orientation*, Atlanta, GA: Doctoral Dissertation, School of Electrical and Computer Engineering, Georgia Institute of Technology, 2005.
- [103] L. Ehrman and A.D. Lanterman, "Target Identification Using Modeled Radar Cross Sections and a Coordinated Flight Model," in *Proceedings of the Third Multi-National Conference on Passive and Covert Radar*, Seattle, WA, 2003.
- [104] L. Ehrman and A.D. Lanterman, "Extended Kalman filter for estimating aircraft orientation from velocity measurements," *IET Radar, Sonar, and Navigation Transactions*, vol. 2, issue 1, 2008, pp.12–16.
- [105] L. Ehrman and A.D. Lanterman, "A robust algorithm for automatic target recognition using precomputed radar cross sections," in *Automated Target Recognition XIV Proc.*, SPIE 5426, Orlando, FL, April 2004.
- [106] E. Greneker and J. Geisheimer, "The Use of Passive Radar for Mapping Lightning Channels in a Thunderstorm," in *Proceedings of the IEEE International Radar Conference*, Huntsville, AL, 2003.
- [107] Kay, S., Xu, C., and Emge, D., "Chemical detection and classification in Raman spectra," in *SPIE Proceedings on Signal and Data Processing of Small Targets*, vol. 6969, 2008.
- [108] Higdon, N.S., Chyba, T.H., Richter, D.A., Ponsardin, P.L., et. al., "Laser interrogation of surface agents (LISA) for chemical agent reconnaissance," in *SPIE Proceedings on Chemical and Biological Sensing III*, vol. 4722, 2002.
- [109] Sedlacek, A.J., Ray, M.D., Higdon, N.S., and Richter, D.A., "Short-range non-contact detection of surface contamination using Raman lidar," in *SPIE Proceedings on Vibrational Spectroscopy-based Sensor Systems*, vol. 4577, 2002.

- [110] T.M. Cover and J.A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [111] P. Howland, "Target Tracking Using Television-Based Bistatic Radar," *IEE Proceedings on Radar Sonar, and Navigation*, vol. 146, pp. 166–174, 1999.
- [112] H. Griffiths and C. Baker, "Passive Coherent Location Radar Systems: Part 1: Performance Prediction," *IEE Proceedings on Radar, Sonar, and Navigation*, vol. 152, pp. 153–159, 2005.
- [113] H. Griffiths and N. Long, "Television-Based Bistatic Radar," *IEE Proceedings Part F*, vol. 133, pp. 649–657, 1986.
- [114] H. Griffiths and C. Baker, "Measurement and Analysis of Ambiguity Functions of Passive Radar Transmissions," in *Proceedings of the IEEE International Radar Conference*, Arlington, VA, 2005.
- [115] H. Lin and A. Ksienski, "Optimum Frequencies for Aircraft Classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 17, pp. 656–665, 1981.
- [116] Y. Lin and A. Ksienski, "Identification of Complex Geometrical Shapes by Means of Low-Frequency Radar Returns," *The Radio and Electronic Engineer*, vol. 46, pp. 472–486, 1976.
- [117] M. Tobias and A.D. Lanterman, "Probability Hypothesis Density-Based Multitarget Tracking with Bistatic Range and Doppler Observations," *IEE Proceedings on Radar, Sonar, and Navigation*, vol. 152, pp. 195–205, 2005.
- [118] B. Bhanu, "Automatic Target Recognition: State of the Art Survey," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 4, pp. 364–379, 1986.

14.11 | PROBLEMS

1. What are the four steps in the unified framework for ATR?
2. What does the following confusion matrix suggest about the feature set used to discriminate Targets 1 and 2?

	Target 1	Target 2
Target 1	0.3	0.7
Target 2	0.2	0.8

3. Given the following confusion matrix, what is the probability of mis-identifying Target 1 as Target 2?

	Target 1	Target 2
Target 1	0.7	0.3
Target 2	0.1	0.9

4. Suppose that Target 1 (from problem 3) can also be deployed in a slightly different configuration, labeled Target 1A. Based on the following confusion matrix, will the existing feature set (used to make the confusion matrix) suffice?

	Target 1	Target 1A	Target 2
Target 1	0.50	0.20	0.30
Target 1A	0.15	0.45	0.40
Target 2	0.03	0.27	0.70

Multitarget, Multisensor Tracking

Lisa M. Ehrman

Chapter Outline

15.1	Review of Tracking Concepts	669
15.2	Multitarget Tracking	677
15.3	Multisensor Tracking	691
15.4	Summary	695
15.5	Further Reading	695
15.6	References	696
15.7	Problems	698

15.1 | REVIEW OF TRACKING CONCEPTS

Once the front end signal processor in the radar has detected targets and converted these detections into measurements (i.e., noise-corrupted observations of the target), the tracker is responsible for assembling the measurements into tracks, each of which includes a state vector and covariance matrix. The first half of this process, which assigns measurements to tracks (and also includes initiation of new tracks and deletion of stale ones), is appropriately termed measurement-to-track association. The latter half of the process is referred to as track filtering. As will be shown, measurement and track covariances are fundamental to both steps.

Organization

Hence Section 15.1.1 gives a brief overview of the covariance matrix, in the context of target tracking. Sections 15.1.2 and 15.1.3 then deliver a brief review of these two processes, thereby providing the requisite background for discussions of multitarget and multisensor tracking in Sections 15.2 and 15.3, respectively.

Key Points

- Target tracking is composed of two main steps: measurement-to-track data association and track filtering.
- The more the measurement covariances for multiple targets overlap, the greater the data association ambiguity. Several techniques are available for addressing this ambiguity, including feature-aided tracking, multiple hypothesis tracking, and cluster tracking.

- Tracking performance suffers when the track filter predictions fail to match the true target dynamics. The IMM estimator and VS-IMM estimator are recommended for addressing target sets with multiple possible modes of target dynamics.
- Multisensor tracking (or sensor fusion) is increasingly common for four primary reasons: necessity, observability, capacity, and robustness.
- Multisensor tracking is not a panacea. It shows great potential, but if executed poorly, it can actually degrade single-sensor tracking performance.

Notation

The following lists many of the variable names from this chapter:

x = state vector

P = covariance matrix

λ = eigenvalue of the covariance matrix

θ = orientation angle of the covariance ellipse

$c_{(i,j)}$ = cost of associating track i and measurement j

c_I = cost of initiating a new track in measurement-to-track data association

G = guard value, or cost of coasting a track in measurement-to-track data association

Λ_{ij} = log-likelihood between track i and measurement j

\tilde{z}_{ij} = innovations, or the difference between the propagated state of track

i (after a measurement conversion) and the state of measurement j

S_{ij} = innovations covariance between track i and measurement j

z_j = state of measurement j

R_j = covariance of measurement j

H_{ij} = measurement matrix, or the Jacobian of the coordinate conversion from the track state space to the measurement state space, linearized around the track state

M_{ij} = Mahalanobis distance between track i and measurement j

F = propagation matrix in the Kalman Filter

K = Kalman gain

\Re_{0j} = SNR of measurement j

\Re_i = Average SNR of track i

\Re_{th} = SNR detection threshold

μ^l is the probability of mode l in the IMM estimator

$\mu^{l,m}$ = probability of having been in mode l at t_{k-1} given that mode m is the best at t_k

$p(l,m)$ = probability of switching from mode l to mode m

\bar{c}_m = probability of being in mode m

Acronyms

CEC	cooperative engagement capability
CRLB	Cramer-Rao lower bound
EKF	extended Kalman filter
GMTI	ground moving target indicator
HOMHT	hypothesis oriented multiple hypothesis tracker
IMM	interacting multiple model estimator
JVC	Jonker-Volgenant-Castanon auction algorithm
LLR	log-likelihood ratio
MHT	multiple hypothesis tracker

pdf	probability density function
pmf	probability mass function
RCS	radar cross section
SNR	signal-to-noise ratio
TOMHT	track oriented multiple hypothesis tracker
TPM	transition probability matrix
VS-IMM	variable structure interacting multiple model estimator

15.1.1 Covariances

The covariance (for real-valued variables) is defined as

$$P = E[(x - \mu)(x - \mu)^T] \quad (15.1)$$

where P is the covariance, x is the random variable, and μ is the random variable's mean [34]. In the context of target tracking, x is the $N \times 1$ track state vector being estimated (which might include position and velocity in a Cartesian coordinate system), and P is the $N \times N$ track covariance matrix, which describes the coupled uncertainties of the elements in the state. The terms on the main diagonal of P are the variances of each term in x , and the off-diagonal terms correspond to the respective cross-covariances.

In target tracking literature, the track covariance is commonly represented as an N -dimensional ellipsoid centered on the track state. As is shown in Figure 15-1 for a 2-D case, the axes of this ellipsoid are equal to the square roots of the eigenvalues, λ_1 and λ_2 , of the covariance matrix, and the orientation of the ellipsoid, θ , is a function of the eigenvectors. This physical representation of the track covariance matrix as an ellipsoid is based upon the assumption that the underlying statistics associated with the track estimation process are Gaussian.

From this assumption, it follows that some percentage of the observations of a track state are contained in the (accurately reported) covariance ellipsoid. Just as 68.3% of the observations of a 1-D Gaussian random variable will be contained within its standard deviation, 68.3% of the observations of a 1-D state are contained within its (degenerate, in the 1-D case) covariance ellipsoid. More generally, the percentage of observations contained in the ellipsoid is a function of the state dimensionality and the number of standard-deviations (or sigmas, for short). Given the assumption of underlying Gaussian statistics, these are related through a chi-squared table.

To recap, the covariance matrix has a physical representation as an N -dimensional ellipsoid, which is equivalent to a contour depicting some probability of containment for the state. Decisions in measurement-to-track data association and track filtering rely heavily on these characterizations of measurement and track uncertainty.

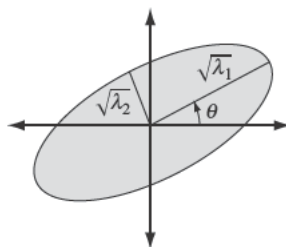


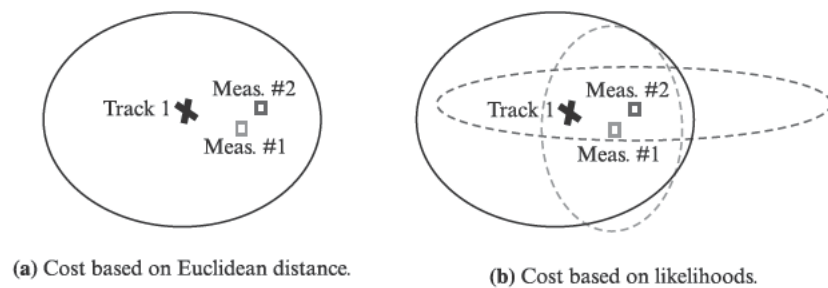
FIGURE 15-1 ■
Representation of
the covariance
matrix as an ellipse.

15.1.2 Measurement-to-track Data Association

Track filtering requires that measurements have been assigned to the track being updated. This process would be trivial if the radar tracked only a single target that always generated detections in an environment free from false alarms. However, this is rarely (if ever) the case.

To illustrate the complications that arise even in seemingly simple measurement-to-track data association problems, consider the case in Figure 15-2a. A tracker is maintaining one track, whose state is represented in the figure by the X and whose covariance ellipse is shown. The tracker then receives two measurements that are both candidates for assignment, represented by the two small squares. It must decide which (if either) to use to update the track and which should become candidates for new track initiation.

FIGURE 15-2 ■ Information leveraged by different measurement-to-track data association cost functions.



Using just the information shown in Figure 15-2a, measurement 1 appears to be the better choice; its state is closer to that of the track. However, a decision based solely on the Euclidean distance between the states would be ill-informed. To illustrate this, consider the notional measurement covariances in Figure 15-2b. The state for track 1 is barely within the containment ellipse for measurement 1; this is analogous to getting a random draw near the tail of the Gaussian distribution. The corresponding likelihood that measurement 1 and track 1 originate from the same object is hence quite low. The state of track 1 is much closer to the center of uncertainty ellipse for measurement 2, which is analogous to getting a random draw near the middle of the Gaussian distribution. As a result, the likelihood that measurement 2 and track 1 originate from the same object is higher, when measurement and track covariances are considered as part of the distance measure.

In real-world radar applications, the number of measurements and tracks may extend into the hundreds, or even thousands. This deluge of data is first typically organized into a measurement-to-track cost matrix, where the N measurements to be assigned appear across the top row and the M existing tracks appear in the left-most column, as in Table 15-1. Note that an extra column (in the far right of the table) is included for tracks

TABLE 15-1 ■ Notional Measurement-to-Track Cost Matrix

	Measurement 1	Measurement 2	...	Measurement N	ϕ
Track 1	$c(1,1)$	$c(1,2)$		$c(1,N)$	G
Track 2	$c(2,1)$	$c(2,2)$		$c(2,N)$	G
...					
Track M	$c(M,1)$	$c(M,2)$		$c(M,N)$	G
ϕ	c_I	c_I		c_I	∞

that are not updated with this particular measurement set. An extra row (on the bottom) is similarly included for measurements that fail to assign to any track. Thus, G is the cost of coasting a track (also referred to as the guard value), while c_I is the cost of initiating a new track.

15.1.2.1 Gating

Before populating the $(1,1)$ through (M,N) entries of the measurement-to-track cost matrix (from Table 15-1), it is prudent to identify infeasible associations. Their costs can immediately be set to an arbitrarily large number, alleviating the computational burdens associated with computing their exact costs and considering them in assignment hypotheses. Given the large number of measurements and tracks that are typical in real-world applications, this step is usually necessary to achieve real-time tracking.

Infeasible assignments are typically found with a series of tests called gates; as a result, this process is known as gating. Coarse gates are applied first, to eliminate potential measurement-track pairs that are grossly mis-matched. Coarse gates can take different forms, with spherical and rectangular gates being common, numerically-efficient choices. Spherical gates keep only those measurement-track pairs whose states are within some Euclidean distance, d_{sphere} , of each other, as described by

$$\sqrt{(px_t - px_m)^2 + (py_t - py_m)^2 + (pz_t - pz_m)^2} < d_{sphere} \quad (15.2)$$

where the track position is (px_t, py_t, pz_t) , and the measurement position is (px_m, py_m, pz_m) . Measurement-track pairs outside this radius are considered infeasible and are ruled out in subsequent parts of the measurement-to-track association process. Rectangular gates apply absolute value tests to each element of the state, thus creating a rectangular acceptance region, rather than spherical one. An example of a rectangular gate for position terms is given by

$$\begin{aligned} |px_t - px_m| &< d_{px} \\ |py_t - py_m| &< d_{py} \\ |pz_t - pz_m| &< d_{pz} \end{aligned} \quad (15.3)$$

where d_{px} , d_{py} , and d_{pz} are the allowable thresholds for the differences in x , y , and z position, respectively.

Having ruled out grossly infeasible measurement-track pairs via coarse, numerically-efficient gates, the next step is to apply fine gates. These are typically elliptical in nature, as they incorporate the measurement and track covariances and assume underlying Gaussian statistics. For this step, it is common practice to compute the full log-likelihood, Λ_{ij} , that measurement j and track i originate from the same object, given by

$$\Lambda_{ij} = -\frac{1}{2} \ln(|2\pi S_{ij}|) - \frac{1}{2} \tilde{z}_{ij}^T S_{ij}^{-1} \tilde{z}_{ij} \quad (15.4)$$

where \tilde{z}_{ij} is the innovations and S_{ij} is its covariance. The innovations is the difference between the measurement state, z_i , and the predicted track state, x_i , which is converted to the measurement coordinate space through some (typically nonlinear) coordinate transformation, $h(\cdot)$. Hence, it is given by

$$\tilde{z}_{ij} = z_j - h(x_i) \quad (15.5)$$

The covariance of the innovations, S_{ij} , is then found with

$$S_{ij} = H_{ij} P_i H_{ij}^T + R_j \quad (15.6)$$

where H_{ij} is the Jacobian of $h(\cdot)$ linearized around the track state, P_i is the covariance of track i , and R_j is the covariance of measurement j . Having computed Λ for each measurement-track pair, the elliptical (fine) gate rules out those pairs whose log-likelihoods exceed some statistical threshold, d_{stat} .

Note that the cost (c_{ij}) typically used to populate the measurement-to-track cost matrix is simply the negative log-likelihood, given by

$$c_{ij} = \frac{1}{2} \ln(|2\pi S_{ij}|) + \frac{1}{2} \tilde{z}_{ij}^T S_{ij}^{-1} \tilde{z}_{ij} \quad (15.7)$$

Although the elliptical gate is not as computationally efficient as the spherical and rectangular gates, it can be re-used later in the track filtering stage.

15.1.2.2 Cost Functions

Before proceeding, it is worth mentioning that other forms of the cost function are sometimes used in real-world systems, with varying degrees of success. Although consensus tends to support use of the full negative log-likelihood, the Mahalanobis distance term, given by

$$M_{ij} = \tilde{z}_{ij}^T S_{ij}^{-1} \tilde{z}_{ij} \quad (15.8)$$

is sometimes treated as a substitute. This can be a perilous choice.

Consider the simple example with two tracks and one measurement. Suppose that the track states are equidistant from the measurement state, and that track 1 reports a perfectly consistent covariance (i.e., one that perfectly reflects the track's error). Further suppose that track 2 over-reports the track error, providing a pessimistic covariance. If the Mahalanobis distance from (15.8) is used, the cost of assigning to the second track will be smaller; the measurement will assign to the track with the larger covariance. If the full negative log-likelihood from (15.7) had been used, the $0.5 \ln(|2\pi S_{ij}|)$ term would have penalized the track with larger covariance, resulting in likely assignment of the measurement to track 1.

Since young tracks often report pessimistic track covariances, they are prone to 'stealing' measurements from older, more-established tracks when the Mahalanobis distance is used for the cost function. Using the full negative log-likelihood mitigates this problem.

15.1.2.3 Assignment Algorithms

Once untenable associations have been eliminated via gating, and the measurement-to-track cost matrix has been populated (preferably with the full negative log-likelihoods) an assignment algorithm is applied to the cost matrix find the best hypothesis. In this context, a hypothesis is a set of assignments. The cost of each hypothesis is the sum of all contributing assignment costs, including costs of coasting or initiating tracks. The best hypothesis is the one with the minimum cost.

Many assignment algorithms are available in the literature, ranging from simple greedy algorithms to more sophisticated, globally-optimal ones. The simplest algorithms start with the first track (i.e., the first row in the measurement-to-track cost matrix), find its minimum-cost assignment, remove that measurement from consideration in subsequent

assignments, and repeat the process sequentially for remaining tracks until all are assigned or coasted. Such approaches are clearly order-dependent and sub-optimal. Given advances in computing and efficient coding techniques, globally-optimal, computationally-efficient auction algorithms, such as the Jonker-Volgenant-Castanon (JVC) algorithm [12, 19, 8], have become popular.

15.1.3 Track Filtering

As measurements are associated into a track over time, the job of the track filter is to update estimates of the track state and covariance. A wide variety of track filters are available in the literature, ranging from simple filters like the alpha-beta filter [10] and Kalman Filter [20] to more sophisticated techniques such as variable-structure interacting multiple model (VS-IMM) estimators [27, 28] and particle filters [39]. Since it is the work-horse in many current applications, the Extended Kalman Filter (EKF) is treated briefly in this section. More sophisticated filters designed to handle ambiguity associated with multi-target scenarios are presented in Section 15.2.

The EKF is composed of two main steps: track propagation and measurement update. At the end of the previous iteration in the EKF (at time t_{k-1}), the state and covariance were brought up-to-date using any measurements assigned to the track at that time. Hence, the state and covariance entering the current iteration of the EKF at time t_k are denoted as $x_{k-1|k-1}$ and $P_{k-1|k-1}$, respectively. The subscripts indicate that the state and covariance were updated at time t_{k-1} given the available data at time t_{k-1} . Before measurement-to-track association can occur at time t_k , the track state and covariance must be propagated to t_k , the current measurement time.

Hence, the first step in the EKF is to propagate the state and covariance forward to t_k using

$$x_{k|k-1} = f_{k-1}(x_{k-1|k-1}) \quad (15.9)$$

and

$$P_{k|k-1} = F_{k-1} P_{k-1|k-1} F_{k-1}^T + G_{k-1} Q_{k-1} G_{k-1}^T \quad (15.10)$$

In this context, f is a (possibly nonlinear) function that propagates the state forward in time. The Jacobian of this (possibly nonlinear) function linearized around the state is denoted by the matrix F_{k-1} , which pre and post-multiplies the covariance matrix. The second term in (15.10), $G_{k-1} Q_{k-1} G_{k-1}^T$, is collectively known as the process noise, and is intended to capture uncertainty about the propagation function, f . If the target always behaved exactly as predicted by f , then this term would be arbitrarily small. A thorough discussion on setting the process noise is given elsewhere [10].

Once the state and covariance are propagated to the measurement time, t_k , measurement-to-track association takes place, as discussed in Section 15.1.2. The second half of the EKF then updates the state and measurement with this new measurement, producing $x_{k|k}$ and $P_{k|k}$, the state and covariance at time t_k given all information available through time t_k . To find the updated state, the EKF essentially takes the weighted average of the track state and measurement state, where the weights are their respective covariances. To accomplish this, it first finds the innovations, or the difference between the measurement state and predicted track state, given by

$$\tilde{z}_k = z_k - h_k(x_{k|k-1}) \quad (15.11)$$

where h is the (possibly nonlinear) coordinate transformation from the track state space to the measurement state space. The covariance of the innovations is found next, via

$$S_k = H_k P_{k|k-1} H_k^T + R_k \quad (15.12)$$

where H_k is the Jacobian of $h(\cdot)$ linearized around the track state, $P_{k|k-1}$ is the predicted track covariance, and R_k is the measurement covariance. Note that (15.11) and (15.12) mimic (15.5) and (15.6). They are repeated here to emphasize the time subscripts, while in (15.5) and (15.6), the subscripts emphasized the particular measurement and track being considered. The Kalman gain then compares the measurement and track covariances, with

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \quad (15.13)$$

The state and covariance are then updated at time t_k with

$$x_{k|k} = x_{k|k-1} + K_k \tilde{z}_k \quad (15.14)$$

and

$$P_{k|k} = [I - K_k H_k] P_{k|k-1} \quad (15.15)$$

Numerical precision errors can sometimes cause the covariance from (15.15) to lose its symmetry or positive definite traits. To mitigate this problem, the equivalent form given by

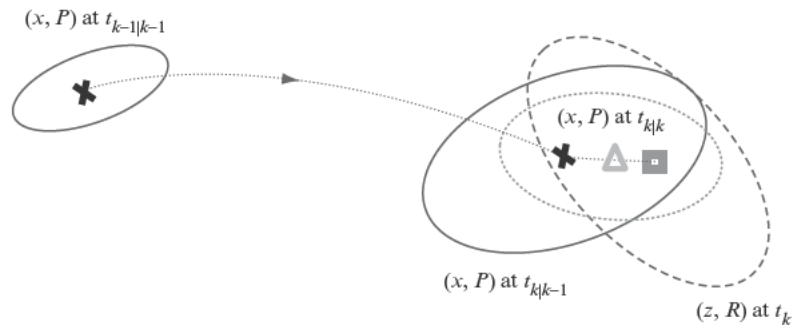
$$P_{k|k} = (I - K_k H_k) P_{k|k-1} (I - K_k H_k)^T + K_k R_k K_k^T \quad (15.16)$$

is often used.

The entire process is depicted in Figure 15-3. The state and covariance at the beginning of the current iteration are depicted by the X and solid-line ellipse on the left side of the figure. They are propagated along the dotted arrow to the X and (inflated, due to process noise) solid-line ellipse on the right side. The assigned measurement and its covariance are depicted with the square and dashed-line ellipse. The final track state and covariance are shown with the triangle and dotted-line ellipse.

The optimality of the Kalman Filter can be proven via multiple techniques (assuming that f and h are linear and the underlying statistics are Gaussian) [20, 32]. In the interest of brevity, these proofs are not repeated here. Rather, a simplified 1-D example is presented, for the purpose of guiding intuition. Suppose for a moment that the state and covariance are both scalars (as would be the case for a 1-D example) and that the measurement and

FIGURE 15-3 ■ Two steps (propagation and measurement-update) comprising the EKF.



state are in the same coordinate systems. In this simplified 1-D example, the Kalman gain reduces to

$$K_k^{(1-D)} = \frac{P_{k|k-1}^{(1-D)}}{P_{k|k-1}^{(1-D)} + R_k^{(1-D)}} \quad (15.17)$$

If the propagated track covariance, $P_{k|k-1}^{(1-D)}$, is much larger than the measurement covariance, $R_k^{(1-D)}$, then the Kalman gain approaches one. Subsequently, the state update simplifies to

$$x_{k|k} = x_{k|k-1} + (z_k - x_{k|k-1}) = z_k \quad (15.18)$$

which effectively sets the updated state equal to the state of the new measurement. This makes intuitive sense. If the measurement's uncertainty is much smaller than that of the propagated track state, then it should be trusted much more heavily. Similarly, if the propagated track covariance is much smaller than the measurement covariance (in this simple 1-D example), then the Kalman gain approaches zero and the measurement does not affect the propagated state and covariance. Again, this makes intuitive sense. If the measurement is much less trustworthy than the propagated state, its impact on the propagated state should be minimal. If the track and measurement covariances are relatively similar in size, then the Kalman gain is between zero and one and the updated state lands along the line connecting the states of the predicted track and measurement.

15.2 | MULTITARGET TRACKING

The more objects observed by the radar, the greater the potential for performance degradation in the tracker. Measurement-to-track association (from Section 15.1.2) is clearly more difficult when many objects are observed in a similar locale. Generally speaking, the more the measurement covariances overlap, the more ambiguous this process becomes. Track filtering (from Section 15.1.3) can also be adversely affected by the presence of many targets. In the standard EKF, the same propagation function (denoted by f in (15.9)) is applied to all tracks. However, the more objects in the scene, the more likely it becomes that some of them will exhibit different target dynamics. A single propagation function may be inadequate to reliably track a large number of (potentially) disparate classes of objects.

Fortunately, target tracking literature is full of sophisticated techniques intended to mitigate these problems. A thorough treatise on all of these techniques would require a separate textbook. Hence, this section is limited to some of the more common techniques for mitigating the complications that multi-target tracking poses to measurement-to-track association and track filtering. With that in mind, Section 15.2.1 covers approaches to multi-target measurement-to-track data association that incorporate features, track clusters, or use multiple-hypothesis tracking. Section 15.2.2 then presents state-of-the-art alternatives for addressing disparate sets of target dynamics, such as VS-IMM estimators.

15.2.1 Measurement-to-track Data Association

Measurement-to-track data association is significantly harder when multiple targets are present. The resulting ambiguity is the rationale for more sophisticated techniques.

Feature-assisted tracking (summarized in Section 15.2.1.1) has the potential to improve measurement-to-track association when the kinematic costs across several elements of the cost matrix are similar; it augments the kinematically-based assignment costs with feature-based costs. Regardless of whether they incorporate features, sophisticated techniques for solving the cost matrices to find optimal assignments, such as multiple hypothesis trackers (MHTs), are sometimes required to maintain sufficiently pure tracks in the presence of dense target scenes. Although these filters incur significant computational costs, they provide a viable option if track purity is required in the context of a dense target scene. The MHT concept is summarized in Section 15.2.1.2. An alternative approach is summarized in Section 15.2.1.3; rather than try to resolve measurement-to-track data association ambiguity, cluster tracking algorithms embrace it and track a set of closely-spaced objects as a group (i.e., a cluster), hence reducing the computational requirements and necessary communications bandwidth.

15.2.1.1 Feature-assisted Tracking

When several elements of the kinematically-based measurement-to-track cost matrix are similar, the ‘correct’ association becomes ambiguous. One technique for resolving this ambiguity is to augment the kinematically-based cost function from (15.7) with a feature-based term. If the kinematics and feature are assumed to be independent (a common assumption to simplify the math, even though it is not really true), then the total cost is simply the product of the kinematic cost and feature cost.

Literature proposing techniques for feature-assisted tracking is abundant [22, 25, 41], some of which focuses on features that are difficult to routinely collect (due to limited radar resources) and much of which focuses on ground moving target indication (GMTI) environments with a limited number of targets. With regards to GMTI, feature-assisted tracking is a useful tool for resolving ambiguities associated with target crossings; comparison of time-invariant features before and after the crossings (e.g., of vehicles at an intersection) can mitigate track swaps. Rather than focusing specifically on GMTI applications with relatively small numbers of targets, or on feature-assisted tracking using features that can rarely be collected *en masse*, this section focuses on recent research applying signal-to-noise ratio (SNR)-assisted tracking to dense target scenes.

Since feature-assisted tracking is often applied in conjunction with an MHT to address scenarios with ambiguous assignment, this discussion uses the full hypothesis cost from MHT literature, given by

$$c_{i,j} = -\ln [P(\Theta^{k,l}|Z^k)] \quad (15.19)$$

where $P(\Theta^{k,l}|Z^k)$ is the probability of hypothesis l (a particular pairing of measurements and tracks) at time k . Prior work [4] derives this to be given by

$$P(\Theta^{k,l}|Z^k) = \gamma \prod_{j=1}^{m(k)} \{f_{i,j}[z_j(k)]\}^{\tau_j} \prod_i [P_{Di}(k)^{\delta_i} (1 - P_{Di}(k))^{1-\delta_i}] P(\Theta^{k-1,s}|Z^{k-1}) \quad (15.20)$$

where $\Theta^{k,l}$ is the joint association of all measurement/track pairings in hypothesis l at t_k , Z^k is the cumulative set of measurements at t_k , $f_{i,j}$ is the probability density function (pdf) of the residual for track i and measurement j , P_{Di} is the probability of detection for track i , τ_j is the indicator function for the assignment of measurement j to any track in hypothesis l , and δ_i is the indicator function for the assignment of any measurement to

track i in hypothesis l . The multiplicative factor, γ , in (15.20) is found with

$$\gamma = \frac{\phi! v! \mu_\phi(\phi) \mu_v(v) V^{-\phi-v}}{(c)m(k)!} \quad (15.21)$$

where ϕ is the number of false alarms at t_k , v is the number of new targets at t_k , $m(k)$ is the number of measurements at t_k , $\mu_\phi(\phi)$ is the probability mass function (pmf) of ϕ , $\mu_v(v)$ is the pmf of v , V is the track gate volume, and c is a normalization constant. Hence, (15.20) provides the complete probability for hypothesis l at time t_k .

If features are not included, then $f_{i,j}$ is equal to the likelihood whose log is shown in (15.4). If the observed SNR is rigorously incorporated into $f_{i,j}$ as a feature [14], then this becomes

$$\begin{aligned} f_{i,j}[z_j(k)]^{\tau_j} &= \exp\left(\Lambda_{i,j}^{\tau_j}\right) p\left(\Re_{0j}|H_1, \Re_i, \Re_{0j} \geq \Re_{th}\right)^{\tau_j} \\ &\quad \times p\left(\Re_{0j}|H_0, \Re_i = 0, \Re_{0j} \geq \Re_{th}\right)^{1-\tau_j} \end{aligned} \quad (15.22)$$

where \Re_{0j} is the observed SNR of measurement j , \Re_{th} is the SNR detection threshold, \Re_i is the (average) SNR of track i , H_1 is the hypothesis that measurement j originates from the target tracked by track i , and H_0 is the hypothesis that measurement j is a false alarm. Hence, if hypothesis l includes assignment of measurement j to track i , then τ_j is equal to one and (15.22) reduces to

$$f_{i,j}[z_j(k)]^{\tau_j} = \exp\left(\Lambda_{i,j}^{\tau_j}\right) p\left(\Re_{0j}|H_1, \Re_i, \Re_{0j} \geq \Re_{th}\right) \quad (15.23)$$

If hypothesis l instead treats measurement j as a false alarm, then τ_j is equal to zero and (15.22) reduces to

$$f_{i,j}[z_j(k)]^{\tau_j} = \exp\left(\Lambda_{i,j}^{\tau_j}\right) p\left(\Re_{0j}|H_0, \Re_i = 0, \Re_{0j} \geq \Re_{th}\right) \quad (15.24)$$

Note that this formulation treats the kinematics and features as though they are independent. This greatly simplifies the math, since finding a closed-form expression for the SNR-based likelihoods as a function of target orientation would be extremely challenging (if not impossible). The resulting multiplication of their likelihoods in (15.22) is hence somewhat *ad hoc* and can allow the feature terms to disproportionately drive assignment results. At a minimum, using the complete hypothesis cost from (15.19) appears to help. Further discussion on this point is found in [14, 15].

If the target's radar cross section (RCS) statistics are well-modeled as Rayleigh-distributed (e.g., Swerling II), then [11, 14] documents that the likelihood that measurement j originates from target i , conditioned on the fact that the detection exceeded the detection threshold, is given by

$$p\left(\Re_{0j}|H_1, \Re_i, \Re_{0j} \geq \Re_{th}\right) = \left(\frac{1}{\Re_i + 1}\right) \exp\left[\frac{\Re_{th} - \Re_{0j}}{\Re_i + 1}\right] \quad (15.25)$$

This likelihood has also been derived for more complex targets (e.g., whose RCS statistics include a Rician-distributed mix of Rayleigh and fixed-amplitude parts) in the literature [15]. Regardless of the assumption of the threat's RCS statistics, the likelihood that the measurement is a false alarm is typically modeled with the exponential distribution,

$$p\left(\Re_{0j}|H_0, \Re_i = 0, \Re_{0j} \geq \Re_{th}\right) = \exp(-\Re_{0j}) \quad (15.26)$$

Hence, by making a few assumptions about the target RCS statistics, the likely number of false alarms, and the likely number of new targets, observed SNR can be incorporated into the rigorous hypothesis costs from (15.19) to potentially improve measurement-to-track data association when the kinematics are ambiguous.

Although this technique has merit, it is not a panacea. The probability of error based on observed SNRs can be derived in closed form for Swerling II targets. This analysis shows that SNR-assisted data association may not be sufficient in the context of Swerling II targets to achieve desired performance unless the observed SNRs are accurate and the differences between the observed SNRs for closely-spaced tracks are substantially different.

To illustrate this point, consider the simple case with two targets in track, the second of which has a larger observed SNR [15]. The SNR of the second target can thus be written as

$$\mathfrak{R}_2 = \mathfrak{R}_1 + \Delta_{\mathfrak{R}} \quad \Delta_{\mathfrak{R}} > 0 \quad (15.27)$$

where \mathfrak{R}_1 is the SNR of the first track and \mathfrak{R}_2 is the SNR of the second track, both of which are in linear units (not dB). Using only the features (e.g., if the kinematic costs were completely ambiguous), an error would occur if the SNR of the measurement from the first object, \mathfrak{R}_{01} , is larger than the observed SNR of the measurement from the second object, \mathfrak{R}_{02} . It follows that the probability of error is given by

$$P_E = \int_{\mathfrak{R}_{th}}^{\infty} \left[\frac{\int_{\mathfrak{R}_{02}}^{\infty} p(\mathfrak{R}_{01}|H_1, \mathfrak{R}_1, \mathfrak{R}_{02}) d\mathfrak{R}_{01}}{\int_{\mathfrak{R}_{th}}^{\infty} p(\mathfrak{R}_{01}|H_1, \mathfrak{R}_1) d\mathfrak{R}_{01}} \right] p(\mathfrak{R}_{02}|H_1, \mathfrak{R}_2, \mathfrak{R}_{02} \geq \mathfrak{R}_{th}) d\mathfrak{R}_{02} \quad (15.28)$$

which reduces to

$$P_E = \frac{1}{2 + \frac{\Delta_{\mathfrak{R}}}{\mathfrak{R}_1 + 1}} \quad (15.29)$$

for Swerling II targets. This makes intuitive sense. If the two targets have identical observed SNRs, then $\Delta_{\mathfrak{R}}$ is equal to zero and the probability of error is 0.5; when the observed SNRs are identical, feature-based association is completely ambiguous.

The degree to which the observed SNRs must be different to achieve a desirably low P_E (assuming Swerling II targets) may come as a surprise. Inspection of (15.29) reveals that even if the second target's observed SNR is twice as large as that of the first (in linear units, not dB), then the probability of error is 0.33. The probability of error (based solely on the SNR term, as in (15.29)) is plotted in Figure 15-4a for Swerling II targets. In this plot, the independent axis is the SNR of the weaker target and the dependent axis is the probability of error from (15.29). A family of curves is presented for varying levels of SNR differences between the weaker and stronger targets. For Swerling II targets, the difference between the weaker and stronger observed SNRs must be large (relative to the SNR of the weaker target) before P_E reaches a desirably low level.

The primary reason for this disappointing result is that the pdfs of observed SNR for Swerling II targets overlap substantially, as shown in [15]. If the target RCS statistics were instead well-modeled as fixed-amplitude, then (15.25) would change and the corresponding pdfs of the observed SNRs would be bell-shaped, rather than exponential-looking. As a result, a satisfactory P_E could be attained via feature-assisted tracking for a much smaller difference in observed SNRs. To illustrate this, Figure 15-4b plots (15.28) assuming fixed-amplitude targets, using the same family of curves as in Figure 15-4a. The P_E

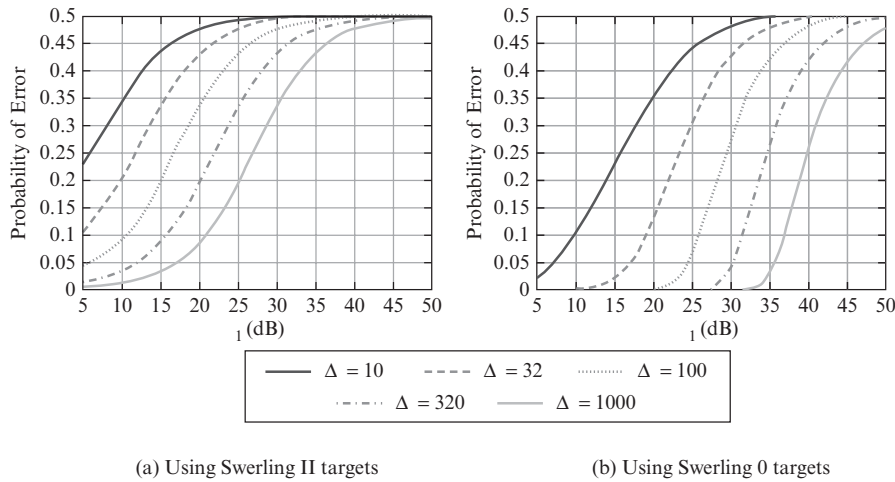


FIGURE 15-4 ■ P_E for SNR-assisted data association.

is clearly much better if the underlying RCS statistics of the targets are well-modeled as fixed-amplitude instead of with the Rayleigh distribution.

While the RCS distribution for a particular object is fundamentally determined by physics, the distribution of the observed SNR can be manipulated by the radar engineer via pulse integration. The details are given in [15], but essentially, the pdf of the observed SNR of a Swerling II target becomes Rician-distributed when multiple pulses are non-coherently integrated. The more pulses that are integrated, the more the Rician pdf of the observed SNR looks like the fixed-amplitude model and the better the P_E that can be achieved via SNR-assisted data association.

For example, Figure 15-5 illustrates that if eight pulses are non-coherently integrated from Rayleigh targets with SNRs of 20 dB and 23 dB (total, after pulse integration), the observed SNR pdfs bear close resemblance to that of a Rician distributions that heavily favor fixed-amplitude pdfs, hence reducing the P_E dramatically. In this example, if the stronger target's SNR is twice that of the weaker one (in linear units); the corresponding P_E drops from 0.33 (with no pulse integration) to 0.08 (when eight pulses are non-coherently integrated). The literature [15] includes techniques for determining the number of pulses that must be integrated to achieve a desired P_E given Swerling II targets.

To recap, if carelessly applied, SNR-assisted data association can actually degrade measurement-to-track association results in multi-target scenarios. This is well-documented in the literature [14, 15]. However, if correctly applied, such feature-assisted

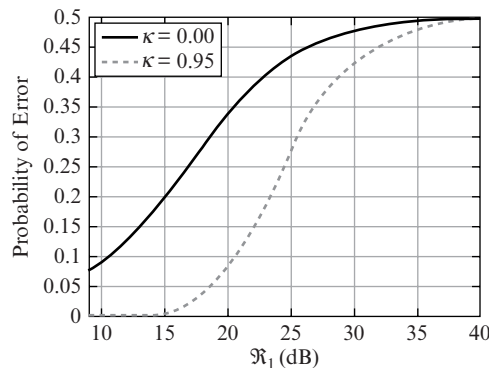


FIGURE 15-5 ■ P_E improvements via pulse integration.

data association has the potential to improve measurement-to-track association when multiple targets are present and the kinematics are ambiguous. Care should be taken not to apply it in an ad hoc manner, but rather, through the rigorous hypothesis cost function from (15.19). By applying it only when the kinematics are ambiguous and when pulse integration is possible, SNR-assisted data association has the potential to resolve the ambiguities brought about by multitarget scenarios.

15.2.1.2 Multiple Hypothesis Tracking

The defining characteristic of MHT is that it attempts to mitigate the effects of transient measurement-to-track association ambiguities by deferring association decisions until the ambiguities resolve. As a result, it is well-equipped to address ambiguous association problems, making it the algorithm of choice for applications that can afford the associated computational complexity and are likely to include dense target scenes.

Two general classes of MHTs have developed in the literature. Reid's seminal paper in 1979 introduced the class of MHTs later labeled as Hypothesis-Oriented MHTs, or HOMHTs [38]. In the mid-1980s, Kurien introduced the Track-Oriented MHT, or TOMHT [23–24] as a more computationally efficient implementation of the MHT. In the interest of completeness, both are introduced in this section.

Hypothesis-Oriented MHT Every time a new set of measurements is collected by the radar, the HOMHT creates an exhaustive list of all possible origins for each [38]. This list includes existing tracks from previous data, new tentative tracks (which would need to be initiated), and false alarms. To illustrate, consider the case in Figure 15-6. Two tracks (T1 and T2) exist based on the previous set of data, and two new measurements (M1 and M2) are collected at time t_k . The first measurement gates with both tracks, while the second measurement gates only with the second track. Hence, the list of possible measurement origins is captured in Table 15-2. Note that T3 refers to the new tentative track that could be initiated on M1, and T4 refers to the new tentative track that could be initiated on M2. The rightmost column in Table 15-2 corresponds to the possibility that the measurements are false alarms.

Next, the list of possible measurement origins (e.g., from Table 15-2) is translated into the exhaustive set of compatible measurement hypotheses. A compatible hypothesis

FIGURE 15-6 ■
Measurement-to-track association example, to illustrate the HOMHT.

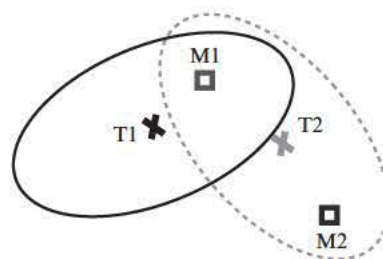


TABLE 15-2 ■ List of Possible Measurement Origins

	T1	T2	T3	T4	ϕ
M1	✓	✓	✓	✗	✓
M2	✗	✓	✗	✓	✓

TABLE 15-3 ■ Exhaustive List of Measurement Hypotheses

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
M1	ϕ	T1	T2	T3	ϕ	T1	T3	ϕ	T1	T2	T3
M2	ϕ	ϕ	ϕ	ϕ	T2	T2	T2	T4	T4	T4	T4

is one whose measurement origins are unique. For example, even though M1 gates with both T1 and T2 in the notional example, the hypothesis that M1 originates from the targets being tracked by both T1 and T2 is incompatible; the two associations are in conflict. In this particular example, eleven compatible hypotheses (H1-H11) are possible, as depicted in Table 15-3. The rows correspond to the potential origins of each measurement, and the columns correspond to the eleven possible hypotheses. For example, the first hypothesis is that both measurements are false alarms. The second is that M1 originates from T1 and that M2 is a false alarm. The eleventh hypothesis is that M1 and M2 are both new targets, which spawn T3 and T4, respectively.

These measurement hypotheses will be carried forward over the next M measurement collections (i.e., scans). Each measurement hypothesis spawns additional sets of measurement hypotheses in each of the next M scans, creating and exponentially growing measurement hypothesis tree. The best measurement hypothesis at time t_k is then chosen at t_{k+M} , based on the best branch in the hypothesis tree at that time. Hence, the HOMHT defers the data association decision for M scans, allowing data association ambiguities to resolve naturally over time.

If computational resources were infinitely available, then the HOMHT could simply carry this entire list of measurement hypotheses forward for M scans, updating and maintaining the resulting tracks in each hypothesis separately. However, given real-world computational constraints, the brute force approach is infeasible (at least for any application with sufficiently dense target scenes to require a MHT). Hence, two main strategies are applied to limit the computations required in the HOMHT. First, the number of measurement hypotheses actually carried forward is typically constrained. This pruning allows only the k -best hypotheses to be carried forward to the next scan, hence limiting the exponential growth of the hypothesis tree over time. Second, rather than updating and maintaining each track separately in each branch of the measurement hypothesis tree, significant savings are often gained by exploiting commonalities of track assignments across hypotheses.

This is accomplished by listing the target hypotheses corresponding to the measurement hypotheses, as in Table 15-4. The target hypothesis list is typically much more compact than the list of compatible measurement hypotheses. Each target (including the new tentative ones) gets a row in this matrix, and each measurement (including the null-symbol, to account for track coasting) gets a column. The checks indicate that the track and measurement are paired in at least one measurement hypothesis. For example, T1 is either coasted (e.g., in H1, H3-H5, H7, H8, H10, H11) or updated with M1 (e.g., in H2, H6, and H9). Hence, even though eleven measurement hypotheses could be carried forward, only two options for T1 must be maintained at this scan: a version where it is coasted and a version where it is updated with M1. Similarly, T2 may be associated with either measurement, or it may be coasted, resulting in checks in all boxes in the T2 row. Hence, three versions of T2 must be maintained at this scan to the possibilities from the set of eleven corresponding measurement hypotheses, as in Table 15-4.

TABLE 15-4 ■ List of Target Hypotheses

	M1	M2	ϕ
T1	✓	✗	✓
T2	✓	✓	✓
T3	✓	✗	✓
T4	✗	✓	✓

TABLE 15-5 ■ Hypothesis Relationship Matrix

	M1	M2	ϕ
T1	H2, H6, H9		H1, H3-H5, H7, H8, H10, H11
T2	H3, H10	H5-H7	H1, H2, H4, H8, H9, H11
T3	H4, H7, H11		H1-H3, H5, H6, H8-H10
T4		H8-H11	H1-H7

A hypothesis relationship matrix is then maintained in the HOMHT to map the versions of each track to each hypothesis. Table 15-5 shows the hypothesis relationship matrix for this example.

Before continuing to the next scan, each measurement hypothesis must be scored. Reid proposes using the Bayesian probability of each hypothesis. Note that (15.20) does not match the formulation in the original paper by Reid, although the intent is the same; Bar-Shalom et al. later extended Reid's original formulation to produce (15.20) [4].

Two main tricks are often used in practice to limit the number of computations required to execute this step. First, a mathematically-equivalent, computationally-preferable, and more numerically-stable log-likelihood ratio (LLR) score is commonly used as a shortcut to computing the hypothesis probabilities. The LLR derivation is explained in depth in numerous sources [8, 9, 42], so will not be repeated here. In short, the LLR takes the natural logarithm of the ratio that the track is valid (i.e., based on target measurements) to the ratio that the track is based entirely on false alarms. The hypothesis score is found by summing the LLRs for each track assignment in the hypothesis. This hypothesis score is then converted to a probability through the approach outlined in [4]. Second, the LLR (which is already more numerically-stable and computationally-convenient than the full hypothesis probability) can be computed as a recursion for each track, minimizing the number of new computations required each time. More details on implementation of this recursion are readily available in [8, 9].

Having scored the measurement hypotheses, the scores can then be used to prune the list, so that only the best are carried forward to the next scan. For example, suppose that H6 had the highest probability, followed by H7 and H5. Further suppose that the remaining eight hypotheses probabilities fell below some allowable threshold. This pruning step would immediately get rid of these eight low-scoring hypotheses; only H5, H6, and H7 would be carried forward to subsequent scans. Note that this can substantially simplify the association problem. In this example, the only three hypotheses remaining after pruning (H5-H7) all assign M2 to T2 and fail to initiate T4. In the next scan, it is thus unnecessary to include T4 or to consider branches from other association alternatives at t_k for T2.

Once the hypothesis scores have all been updated and the low-scoring ones have been pruned at t_k , the next step is retro-active pruning. The best hypothesis (e.g., H6) is identified, and all branches in the tree M scans back (at t_{k-M}) that don't lead to it are

removed. In doing so, the HOMHT produces a definitive association answer M scans after the data are collected, using all the subsequent hypotheses (that exceeded some minimum probability, in practice) created since. The entire process repeats with every new set of measurements that are collected, so that at t_{k+M} , the association answer for t_k will be definitively determined. By deferring the association decision by M scans, the HOMHT allows changes in target motion over time to resolve the previous association ambiguity.

Track-Oriented MHT Recall that the number of possible hypotheses is typically orders of magnitude larger than the number of targets. The TOMHT capitalizes on this fact by carrying tracks (not hypotheses) forward from one scan to the next. As a result of the significant computational savings, the TOMHT is more commonly used and cited in the literature.

When a new set of measurements is collected, the TOMHT begins by independently forming possible associations to tracks. Uniqueness isn't imposed yet in the algorithm. Rather, multiple tracks can be updated with the same measurement at this stage.

The LLR is then computed for each track, using the recursion from [8]. The LLRs for all potential tracks (which need not form compatible sets, or hypotheses yet) are then compared with a threshold. Those that fail this threshold are immediately pruned.

The remaining association candidates are clustered into groups linked by common observations. For example, if measurement 1 could potentially associate with tracks 1, 2, or 3, then these three tracks would be grouped together in a cluster. Furthermore, not all tracks in the cluster have to share a common observation; they could be linked by another track that shares common observations with both. Bear in mind that the intent of this clustering step is not to form hypotheses, or compatible sets of associations—That happens in the next step. Rather, the intent is to divide the association candidates into separable groups so that a series of small association problems can be solved in lieu of one large problem. Since the computational complexity eventually required to solve the problems grows exponentially as a function of the number of tracks in the cluster, it is advantageous to break the associations into as many small clusters as possible.

Finally, once poor association candidates have been pruned out based on their LLRs and once the remaining association candidates have been segregated into separable clusters, the TOMHT forms hypotheses. The candidate associations, along with their updated LLRs and compatibility constraints, serve as inputs for this multi-frame association (MFA) problem, where the frames correspond to the previous scans of data that have been collected. The most common technique for efficiently solving the MFA problem is Lagrangian relaxation [35, 36, 37], the goal of which is to maximize the hypothesis score under the constraint that no tracks in the hypothesis share observations (i.e., are incompatible with each other). Hence, the hypothesis score serves as the objective function, and the constraints are replaced with well-chosen Lagrangian multipliers, allowing efficient creation of hypotheses.

Pruning then occurs again, in the interest of limiting computational complexity. The recently-created hypothesis scores are equal to the sum of their contributing track scores. Any hypotheses or tracks that score too low at time t_k are dropped, to reduce the computational burden. Associations M scans back are also finalized based on the current hypotheses. Only the t_{k-M} associations in the same branch as the current best hypothesis are maintained. Hence, the TOMHT also allows associations to be deferred until additional data is available, so that subsequent target motion has a chance to resolve the ambiguities.

Finally, the measurement update step (e.g., described by (15.11)–(15.16) for an EKF) is performed for the tracks that received updates in the surviving hypotheses. Redundant tracks (i.e., duplicate tracks on the same object) are also sought at this stage, so that they can be deleted. The process then repeats the next time that measurements are collected.

Summary Regardless of the particular implementation, the key advantage of the MHT is that it allows data association ambiguities to resolve naturally over time, often producing tracks that are more accurate and pure than they would be using on a single-hypothesis tracker. The MHT does not dictate a particular cost function of assignment algorithm (although the cost in (15.20) and the JVC assignment algorithm are popular choices); it simply populates and solves these cost matrices using data collected over a longer period of time than single-hypothesis trackers. This translates into significant computational requirements, suggesting that MHTs be used only when the application includes sufficient association ambiguity to require them.

15.2.1.3 Cluster Tracking

The MHT (from Section 2.1.2) defers data association decisions, in the hope that additional scans of data will mitigate the ambiguity. An alternative approach when faced with a dense target scene is to capitulate. Rather than attempting to resolve these ambiguities, cluster tracking (presented in this section) leverages the fact that multiple targets are closely-spaced and tracks them as a group (i.e., a cluster), thereby reducing the computational demands and requisite communications bandwidth [43]. This may be preferable to an MHT if the data association ambiguity isn't likely to resolve quickly or if the computations required by the MHT cannot be accommodated. It could also be used in tandem with a MHT (e.g., by arranging hypotheses that include cluster tracks); although often treated this way in the literature, they need not be mutually exclusive.

Approaches to cluster tracking documented in the literature fall into three main categories [43]. On one extreme are the papers that propose creating and maintaining only cluster track information. Tracks on individual objects in the cluster are not created or maintained. On the other end of the spectrum are cluster tracking algorithms that create and maintain individual tracks, using the cluster information only as a supplement for those individual tracks. In between these two extremes are approaches that create and maintain both cluster and individual tracks; the emphasis in these more-moderate approaches to cluster tracking is on the cluster track, and individual tracks supplement the cluster track, rather than the other way around.

While it appears to have potential for addressing dense target scenes, cluster tracking is fraught with challenges [43]. Cluster management algorithms are required to effectively initiate, merge, split, and delete clusters as sets of closely-spaced (and possibly unresolved) targets shift over time. Addressing objects that enter and exit a cluster requires similar deftness. Clusters of objects may also cross or overlap at times, complicating matters. Nevertheless, the potential computational and communications bandwidth savings render cluster tracking an area ripe for further research.

15.2.2 Modeling Target Dynamics

In the standard EKF, the same propagation function is applied to all tracks, to bring the states and covariances from the previous measurement time forward to the new measurement time. The more objects in track, the more likely it becomes that one (or more) of them will have dynamics not well-modeled by the propagator and process noise. When this

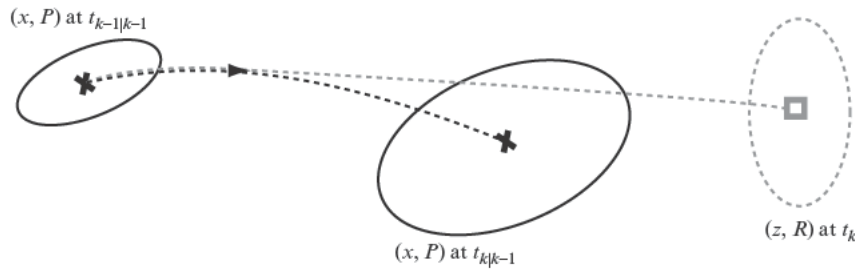


FIGURE 15-7 ■
Illustration of how maneuvering targets can undermine measurement-to-track association.

mismatch occurs, that predicted track's covariance becomes optimistic (e.g., it under-reports the true error). This, in turn, skews the trade-off between the track covariance and measurement covariance in the Kalman gain (in (15.13)), leading the updated state (from (15.14)) to favor the predicted state more than is appropriate. As a result, the track error increases.

If the mis-match is sufficiently severe, it can even prevent measurements (i.e., from a maneuvering target) from associating with the track (i.e., based on a non-maneuvering propagator), resulting in track coasting; over time, this can lead to track churning as tracks on the maneuvering object are sequentially initiated and dropped. This is depicted in Figure 15-7. The state and covariance at the beginning of the current iteration are depicted by the 'x' and solid-line ellipse on the left side of the figure. They are propagated (assuming no maneuver) along the dotted arrow to the 'x' and solid-line ellipse on the right side. The object's measurement and covariance are depicted with the square and dashed-line ellipse. If the target maneuvers, then the measurement may be sufficiently far from the predicted state (due to this mis-match in modeling the target dynamics) that the predicted track covariance and measurement covariance do not overlap; this can prevent data association and result (over time) in a dropped track. If the mis-match continues (i.e., if the target dynamics continue to be mis-matched with the filter), then the process of track initiation and deletion will repeat, resulting in track churn.

Techniques for mitigating this problem have steadily evolved, as depicted in Figure 15-8. The first batch of techniques to be published adapted the single tracker's parameters (e.g., its process noise) to facilitate track maintenance in the presence of a mis-match between the target dynamics and the filter's propagation function [1]. In this adaptive filter approach, the innovations term and its covariance (from (15.11) and (15.12)) are monitored. Once they exceed some pre-determined threshold, the filter parameters are altered to improve the odds of maintaining the track. A common approach is to increase the process noise substantially when a mis-match is detected, and to later reduce it once the mis-match ends. This adaptive filter approach is depicted on the far left side of Figure 15-8.

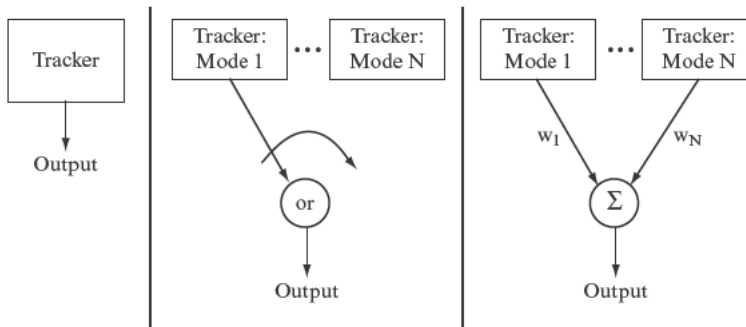


FIGURE 15-8 ■
Evolution of techniques for tracking target with multiple modes of dynamics.

Adaptive filters are fundamentally limited. The process noise is increased (to accommodate track maintenance in the presence of mis-matched target dynamics) at the expense of track accuracy. These limitations drove the literature toward multiple model approaches, as depicted in the middle of Figure 15-8. These multiple model schemes run multiple trackers in parallel; the goal is to provide a tracker for each likely mode of target dynamics. At any given time, the mode yielding the smallest innovations is selected. More advanced versions of this algorithm incorporate hysteresis, so the selection doesn't bounce between modes on each sequential update.

While multiple model schemes provide an improvement over adaptive filters (for many applications), they are not ideal. The biggest complaint lodged against these schemes is that the overall filter output can contain sharp discontinuities over time as the selected mode jumps from one tracker to another. Covariance consistency (over time) can also suffer in these schemes.

Hence, interacting multiple model (IMM) estimators were eventually developed as a way of addressing the mis-match between target dynamics and the track filter [31]. This idea (depicted on the right side of Figure 15-8) is similar to the multiple mode schemes, but rather than selecting the state and covariance from a single filter, the results of multiple filters are blended together. The weights, w_i , for each track filter correspond to the probability that each mode best represents the target dynamics. By incorporating the results from all the parallel filters, the IMM estimator is able to provide more consistent covariances and strike a better balance between track maintenance and track accuracy. Section 15.2.2.1 further discusses the IMM estimator; Section 15.2.2.2 briefly presents a variation on this scheme that addresses one of its key shortcomings.

15.2.2.1 Interacting Multiple Model (IMM) Estimators

Figure 15-9 gives a more detailed view of the IMM estimator [31]. First, previous states and covariances from all contributing modes are mixed together, creating the mixed states and covariances that will be propagated to the current time in the filters associated with each mode. The mixed state and covariance for each mode, l , is accomplished via

$$x_{k-1|k-1}^m = \sum_{l=1}^N \tilde{x}_{k-1|k-1}^l \mu^{l,m} \quad (15.19)$$

and

$$P_{k-1|k-1}^m = \sum_{l=1}^N \mu^{l,m} \left[\tilde{P}_{k-1|k-1}^l + \left(\tilde{x}_{k-1|k-1}^l - x_{k-1|k-1}^m \right) \left(\tilde{x}_{k-1|k-1}^l - x_{k-1|k-1}^m \right)^T \right] \quad (15.20)$$

where $\tilde{x}_{k-1|k-1}^l$ refers to the updated state of mode l at t_{k-1} , projected into the state space of the minimum-order filter¹, $\tilde{P}_{k-1|k-1}^l$ refers to its covariance (also projected into the state space of the minimum-order filter), and $\mu^{l,m}$ is the probability of having been in mode l

¹The modes in the VS-IMM may include states with different dimensions. For example, a constant-velocity mode may model 6-D states (including position and velocity in x, y, and z), while a constant-acceleration mode may model 9-D states (including position, velocity, and acceleration in x, y, and z). When mixing the states and covariances together, the larger-dimension states must be reduced to the size of the smaller-dimension ones. For example, acceleration is dropped from the 9-D state so that the state can be added to the 6-D state. The projected states and covariances are denoted as \tilde{x} and \tilde{P} , respectively.

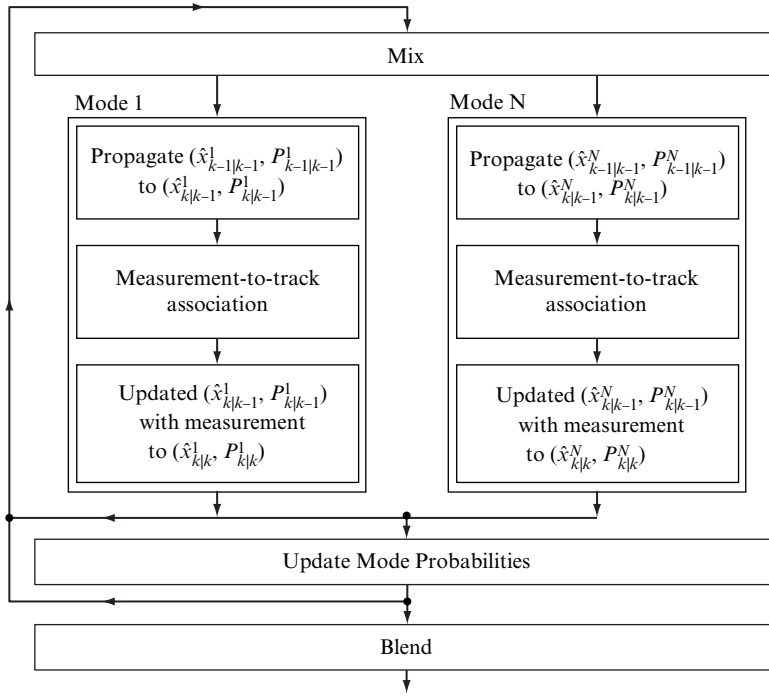


FIGURE 15-9 ■
Steps in the IMM
estimator.

at t_{k-1} given that mode m is the best at t_k . This probability is given by

$$\mu^{l,m} = \frac{p(l,m) \mu^l}{\bar{c}_m} \quad (15.21)$$

where $p(l,m)$ is the probability of switching from mode l to mode m (which is defined prior to the engagement and remains static), μ^l is the probability of mode l at t_{k-1} , and \bar{c}_m is the probability of being in mode m at t_k , given by

$$\bar{c}_m = \sum_{l=1}^N p(l,m) \mu^l \quad (15.22)$$

If the measurement time interval varies with sensors, then the transition probability matrix (TPM) $\{p(l,m)\}$ can vary with time and can be calculated using the algorithm in [30].

The mixed states and covariances are then propagated to the current time in each filter, using (15.9) and (15.10). Note that each of the N filters may have different propagation functions and process noise.

Measurement-to-track association comes next. This step poses a challenge. If each parallel filter independently conducted measurement-to-track association, they could arrive at different answers and begin to represent different objects. Since the goal is to assign the best measurement to this track, based on all the filter representations of it, the likelihoods from (15.4) are combined in a weighted average via

$$\Lambda_{ij} = \sum_{l=1}^N \Lambda_{ij}^l \mu^l \quad (15.23)$$

where the weights are the mode probabilities [17]. The measurement-to-track cost matrix is then populated as a function of these combined likelihoods. As a result, each track is

still represented in the cost matrix with a single row, regardless of the number of modes running in parallel in the IMM estimator.

Once a measurement is assigned to the track, each of the filters in the IMM estimator is updated using the measurement via (15.11) through (15.16). The mode probabilities are also updated at this point, based on the measurement-to-track likelihoods, via

$$\mu^l = \frac{\Lambda^l \bar{c}_l}{\bar{c}} \quad (15.24)$$

where \bar{c}_l comes from (15.22) and \bar{c} is given by

$$\bar{c} = \sum_{l=1}^N \Lambda^l \bar{c}_l \quad (15.25)$$

The updated states and covariances from each filter then become inputs to the mixing function during the next IMM cycle. Should a single estimate be needed from the IMM, it comes from blending the various mode states and covariances together via

$$\hat{x}_k = \sum_{l=1}^N \tilde{x}_{k|k}^l \mu^l \quad (15.26)$$

and

$$P_k = \sum_{l=1}^N \mu^l \left[\tilde{P}_{k|k}^l + \left(\tilde{x}_{k|k}^l - \hat{x}_k \right) \left(\tilde{x}_{k|k}^l - \hat{x}_k \right)^T \right] \quad (15.27)$$

To ensure good track accuracy across the spectrum of possible target behaviors, the IMM designer may be tempted to include a large number of modes. Increasing the number of modes is not always advantageous, though. Each mode influences the track state during the mixing and blending steps. Even though the degree of influence for a given (poorly-matching) mode may be small, the cumulative influence of several poorly-matching modes may be significant enough to noticeably degrade track accuracy. As a result, the IMM designer is faced with a tradeoff between overall track accuracy and the ability to maintain tracks during different dynamic modes. As a rule of thumb, IMM estimators are typically designed with as few modes as possible to achieve adequate track maintenance. They often include filters precisely tuned to the small set of most-likely target dynamics, as well as one generic mode intended to preserve the track when the target dynamics don't match any of the more focused filters.

15.2.3 VS-IMM Estimators

The VS-IMM estimator mitigates the negative impact of this tradeoff by allowing the set of candidate modes to vary adaptively for each track over time [26]. The equations from Section 15.2.2.1 still apply to the VS-IMM. The only difference is that the l modes used at any point in time for a given track are based on that track's current dynamics. As a result, more filters are available to accommodate a broad range of target dynamics. The least-likely filters do not contribute to the mixed or blended states if they are not part of the currently selected set for that track. Given the large number of modes needed to characterize realistic target dynamics in some applications, use of the VS-IMM can be necessary to achieve both satisfactory track accuracy and track maintenance.

15.3 | MULTISENSOR TRACKING

Section 15.2 focuses on the challenges in measurement-to-track data association and track filtering that arise in multitarget scenarios, as well as techniques for mitigating these challenges. This section shifts the focus to the challenges (and mitigating techniques) associated with multisensor tracking, also often termed ‘sensor fusion’.

Sensor fusion is fundamental to numerous radar applications for several reasons, including necessity, observability, capacity, and robustness. This section discusses the rationale for fusion, the two main architectures, and challenges associated with both.

Necessity is the most obvious reason for sensor fusion. Simply put, the field of regard of a radar may be smaller than the area of interest for a given application. When this occurs, multiple sensors are necessary to obtain required coverage. Hence, sensor fusion is often necessary to allow adequate coverage of a particular region.

Even if the radar’s field of regard is sufficiently large, the radar may still have a limited ability to observe features of interest for objects within this region. Suppose, for example, that a radar must detect a small target maneuver. If the maneuver occurs orthogonally to the radar’s location, then the radar’s angle measurements may not be sufficiently accurate to detect it, rendering the cross-range maneuver unobservable. However, that same small maneuver may be observable to another radar at a different location. If spatially-distributed radars are networked, they provide spatial diversity and improved observability to a variety of object features.

The potential for large numbers of targets is inherent in some radar applications. In fact, the number of targets may exceed the (hardware or software) capacity of a single radar. If a single radar cannot adequately service all required functions (e.g., acquisition, tracking, discrimination) for all potential targets, then the obvious solution is to deploy a networked suite of radars. If data are properly associated and fused, the networked system can address a larger capacity of threats than would be possible with any single contributing radar.

Finally, networking and fusion produces a more robust system. A single radar may malfunction or be damaged by an adversary. If radars are intelligently networked, the failure of a single one does not lead to utter failure; the remaining radars can still perform the mission, albeit at a somewhat degraded level (since fewer total observations are available).

15.3.1 Sensor Fusion Architectures

Sensor fusion architectures are fundamentally defined by the type of data (i.e., measurements versus tracks) being fused. Variations are possible (and five options are discussed in [7]), but the two most common fusion alternatives are measurement-level fusion and track-level fusion.

Measurement-level fusion (also sometimes referred to as centralized fusion in academic literature [7]), in which all measurements from all sensors are transmitted and associated to create system tracks, is theoretically the best choice, since it makes use of all available data. However, constraints imposed by communications systems often render this impractical. Consider, for example, a notional set of radars that collect measurements on thousands of objects several times per second. Transmitting all of these measurements from each radar to the rest of the network would require far more bandwidth than is typically available.

Track-level fusion (also sometimes referred to as decentralized fusion in academic literature [7]), is motivated by this constraint. By performing measurement-to-track assignment locally (within each sensor) and only periodically sharing the track states and covariances, the track-level fusion architecture can afford much less frequent transmission of data.

Sensor fusion is required (based on necessity, observability, capacity, and robustness) for many radar applications, but poorly-executed fusion can actually *degrade* results relative to single-radar performance. Both sensor fusion architectures (measurement-level and track-level) are fraught with challenges. Section 15.3.2 provides an overview of the challenges unique to measurement-level fusion. Section 15.3.3 then gives an overview of those unique to track-level fusion, Section 15.3.4 covers the challenges common to both architectures.

15.3.2 Measurement-level Fusion and Associated Challenges

Communication constraints pose the most obvious challenge for measurement-level fusion. Assuming that the sensors are capable of tracking a multitude of targets at high update rates, the sheer volume of measurements that must be transmitted via the communication network can be overwhelming. On-going research is being conducted to improve communication bandwidth and reliability, as well as to adaptively adjust the amount of data that must be transmitted (e.g., by creating pseudo-measurements [7] from a set of measurements).

Measurement-level fusion architectures can also be susceptible to missing, late, or out-of-sequence measurements. This is particularly problematic in architectures like the Cooperative Engagement Capability (CEC), where all sensors receive all measurements from all other sensors and rely on identical fusion and tracking algorithms to maintain synchronized track pictures. A body of work exists on these topics in academic literature [3, 6]; close attention to implementation details is required to make these techniques work in practice.

While sensor biases pose challenges in both fusion architectures, they manifest themselves uniquely in each. In the measurement-fusion paradigm, the measurements must be transformed from local sensor coordinates (e.g., in sine space or range/azimuth/elevation) into Cartesian coordinates (e.g., Earth-centered, Earth-fixed) so that they can be fused with measurements from other sensors at disparate locations. To be sufficiently accurate, this conversion requires knowledge of internal sensor biases from the latest round of calibration. Two main options for addressing this problem are available, but both consume precious communications bandwidth. Either the latest calibration results must be conveyed from the sensors to the fusion center, or the coordinate conversion must occur locally (within the sensors) prior to transmission. The former option requires periodic transmission of (a potentially large amount of) calibration data, while the latter requires frequent transmission of larger measurements (e.g., 6×1 state and 21 unique covariance terms, instead of a 3×1 state and 6 unique covariance terms).

15.3.3 Track-level Fusion and Associated Challenges

Certain challenges are unique to track-level fusion. In particular, prior work [5] demonstrates that estimation errors reported by two sensors on the same target are not independent. Rather, dependence arises due to common process noise used in the respective track

filters [5]. Unaccounted for, this dependence can degrade track-to-track association by distorting the track-to-track costs. This dependence manifests itself in the cross-correlation term, which is a function of each filter's Kalman gain, process noise, and measurement matrix. Although the expression for this term is known [5] and could be accounted for in the track-to-track association costs (in theory), this is practically constrained by the fact that it would require sensors to transmit the Kalman gain matrix, process noise matrix, and measurement matrix with every track state and covariance report. Since communication systems cannot typically accommodate this, a variety of *ad hoc* track-to-track fusion algorithms (including naïve fusion, Chernoff fusion, Shannon fusion, Bhattacharyya fusion, channel fusion, and minimum mean square error fusion) have been developed to address this dependence. The larger the process noise used in the filters, the more failing to account for this dependence can corrupt track-to-track association.

Short-lived sensor tracks (i.e., tracks that end before they sufficiently settle) also pose a unique challenge for track-level fusion schemes. These can occur for a variety of reasons. If the sensor acquires a target shortly before the target leaves its field of view, then the track may not have time to settle before it can no longer be updated by the sensor. Targets with low RCS can similarly be dropped (due to missed detections) before settling. Extremely dense target scenes that foil measurement-to-track association can also result in a series of short-lived, inaccurate tracks. Maneuvers beyond those anticipated by the sensor track filter can also result in tracks that break before they settle. Whatever the cause, short-lived tracks can wreak havoc in track-level fusion. Since they are not settled, they may associate easily with a variety of tracks from other sensors, thereby corrupting system-track feature sets, upon which discrimination or subsequent targeting decisions may be based.

Sensor biases are particularly problematic in track-level fusion architectures. One primary goal of track-level fusion is the reduction of track uncertainty through fusion of tracks from different viewing angles. The fused covariance (which quantifies this uncertainty) typically shrinks such that it is on par with the intersection of the contributing covariances. However, when the contributing sensors fail to properly account for biases, they under-report the uncertainty. As a result, the fused track state may be pulled away from truth by a biased sensor, and the fused covariance may be reduced to the point that the true object in track no longer falls within the reported uncertainty region. This is depicted in Figure 15-10. The left side of the figure depicts two accurate covariances in the solid and dotted lines, respectively; the fused track covariance (shown with the dashed line) is roughly equal to the intersection of the two contributing regions, hence reducing uncertainty. The right side of the figure shows the degradation that occurs when the second sensor reports a biased track. The dotted-line ellipse is centered in the wrong location, resulting in a fused covariance ellipse (dashed-line) that no longer contains the true object reports (dot-dashed-line). This type of degradation is particularly disruptive in applications that require acute situational awareness or fire control.

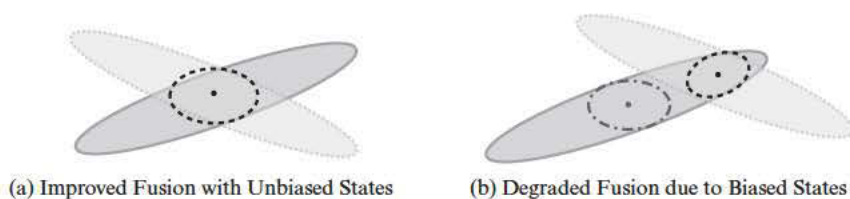


FIGURE 15-10 ■
Example where
biases degrade the
fused result.

15.3.4 Sensor Fusion Challenges Common to Both Paradigms

The aforementioned challenges are unique to measurement-level or track-level fusion, but several additional challenges are common to both fusion architectures. As was previously discussed, sensor fusion is fundamentally a function of the contributing reports' uncertainties. Measurement-level fusion is heavily influenced by reports with small measurement covariances, and track-level fusion is similarly influenced by reports with small track covariances. One significant challenge is that sensors and trackers are often poor judges of their uncertainties, upon which fusion depends. They typically account for known biases (via calibration), but these address only the known-unknowns. Calibration cannot address the host of unknown factors that increase error, or the unknown-unknowns. Hence, radars frequently under-report errors. Techniques for improved bias estimation are widely available in the literature, but application of 'fudge factors' that inflate reported covariances are often used in practice, even though they are sub-optimal.

Depending on the application, the sensors whose results are being fused may be deployed over a large geographical space. If this is the case, then gaps in sensor coverage pose another challenge for both fusion paradigms. Situational awareness (a common goal for multi-sensor fusion systems) mandates that the system track set (i.e., the result of fusion in either scheme) include one track per reported object. If the objects all exit the coverage area for one set of sensors before entering the coverage region of another, this implies that both sets of tracks must be fused to create the desired single system track per object. This, in turn, requires the states and covariances to be propagated to a common reference time. Depending on the sensor laydown, this may translate into tens or hundreds of seconds of propagation. The challenges compound. Fusion of perfectly unbiased, fully-settled tracks with perfectly consistent covariances may not be terribly degraded by long propagation times, but long propagation times will severely degrade fusion results if the tracks are biased, are short-lived, or have inconsistent covariances. Propagation schemes that identify the optimal time frame in which to conduct association may mitigate this challenge [16].

The deployment of disparate radars further compounds these problems. Radars with different sensitivities, range resolution, viewing aspects, and coverage are unlikely to report on identical objects from the threat complexes. In fact, the subsets of objects that each report may not overlap significantly. Notionally, one sensor with poor sensitivity and resolution may report a handful of tracks from a given scene, while another with excellent sensitivity and resolution may report hundreds of tracks on the same set of objects. Association and fusion become extremely challenging in this context, sometimes referred to as 'subset conditions', since the subsets of objects being tracked by the different radars may not substantially overlap. Attempting bias estimation and removal in this context (if divorced from physics-based rationale) can be treacherous, as numerous, non-unique biases may cause the small set of tracks to align well with subsets of the larger one, at least for short periods of time.

Whether the contributing radars report discrimination feature observations (as would be likely in a measurement-level fusion scheme) or track-based discrimination answers (in the track-fusion paradigm), fusing discrimination answers from multiple sensors is extremely problematic. Just as disparate sensors at differing deployment locales will report different numbers of tracks on the same complex, they may also collect different features (with different levels of accuracy) with different intended uses. While work has begun in this area, it is ripe for further research.

15.4 | SUMMARY

Target tracking consists of two main steps: measurement-to-track data association and track filtering. When multiple targets are present in proximity, both steps are more prone to errors. The more the measurement covariances for multiple targets overlap, the greater the data association ambiguity. This chapter presented three sets of techniques for achieving good performance in the face of this ambiguity. One option is to incorporate features into the measurement-to-track cost matrix. This can be a reasonable approach, if the chosen feature is readily and accurately observable and if it segregates the target set. Regardless of the chosen cost function, another approach to resolving measurement-to-track ambiguity is to defer the decision until additional scans of measurements have been collected. The MHT uses this approach, which works well in cases where the ambiguity is likely to resolve over time. In some cases, the targets may be unresolved or very closely-spaced for long periods of time, necessitating cluster tracking.

Track filtering can also be more challenging in multitarget scenarios; the more objects in track, the more likely it may be that the set of targets will include some with disparate dynamics. Track accuracy, and even track maintenance, suffer when the track filter predictions fail to match the true target dynamics. The IMM estimator is a common approach for coping with this problem, and can address both a single object that has multiple potential dynamic modes and a set of objects with potentially disparate dynamic modes. While the IMM estimator provides a good tradeoff between track accuracy and track maintenance, its performance can suffer when a large number of modes are required to adequately address the observed target classes. The VS-IMM combats this problem by allowing the number of modes in the IMM to adapt over time for each track.

Multisensor tracking (or sensor fusion) is increasingly common, for a number of reasons. Many applications require coverage of a region larger than that which can be viewed by a single sensor; hence, sensor fusion arises out of necessity to cover the entire area of interest. Fusion of data from spatially-distributed sensors also offers improved target observability. The capacity of the system also increases via multisensor tracking; the more radars in the system, the more objects that it can track. Finally, multisensor tracking improves system robustness, as the system can withstand loss (due to reliability failures of malicious adversary actions) of a subset of its sensors.

A number of multisensor architectures are possible, with measurement-level fusion and track-level fusion being the most common choices. Certain challenges are inherent to each and must be addressed to ensure that the sensor fusion result is actually better than that of the individual contributing sensors.

15.5 | FURTHER READING

The literature on multitarget, multisensor tracking is vast and rapidly evolving. With this in mind, the *Journal for Advances in Information Fusion* is recommended for the reader who wishes to stay abreast of advances in sensor fusion. This excellent, free, on-line resource is updated twice each year and contains peer-reviewed, journal-quality papers on sensor fusion.

While far from exhaustive, the following references are also recommended for the reader who is interested in multitarget tracking metrics, performance prediction, and non-Gaussian/nonlinear filtering applications.

Metrics for multitarget tracking applications are often difficult to implement in real-world settings (as opposed to simulations), where the targets being tracked may be unresolved, hence making the true origin of each measurement somewhat ambiguous. Kamen developed an approach for multitarget tracking using symmetric measurement equations [21] with potential application to metrics, and Daum built on this to derive the Cramer-Rao Lower Bound (CRLB) for multitarget tracking [13]. For a discussion of the difficulties of scoring multitarget tracking algorithms, see [33, 40].

In the interest of brevity, performance prediction was outside the scope of this chapter. However, it is a relevant topic for those involved in algorithm development or system acquisition. Performance prediction literature includes papers focused on the measurement-to-track association step [2] as well as the track filtering step [27, 28].

The multitarget, multisensor tracking literature also includes sophisticated techniques for coping with highly-nonlinear or non-Gaussian problem sets. For example, unscented Kalman Filters [18] have been demonstrated to be superior to extended Kalman Filters when the propagation function and/or the measurement conversion function are highly-nonlinear. Particle filters [39] have been developed to accommodate applications in which the underlying estimation statistics are non-Gaussian. Mahler [29] has derived the probabilistic hypothesis density algorithm using finite-set statistics, to simultaneously estimate the number of targets present and estimate their states.

15.6 | REFERENCES

- [1] Alouani, A.T., Xia, P., Rice, T.R., and Blair, W.D., "A Two-Stage Kalman Estimator for State Estimation in the Presence of Random Bias and for Tracking Maneuvering Targets," *Proceedings of the 30th IEEE Conference on Decision and Control*, pp. 2059–2062, 1991.
- [2] Areta, J., Bar-Shalom, Y., and Rothrock, R., "Misassociation Probability in M2TA and T2TA," *Journal for Advances in Information Fusion*, vol. 2, no.2, pp. 113–117, 2007.
- [3] Bar-Shalom, Y., "Update with Out-of-Sequence Measurements in Tracking: Exact Solution," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, pp. 769–777, 2002.
- [4] Bar-Shalom, Y., Blackman, S., and Fitzgerald, R., "The Dimensionless Score Function for Multiple Hypothesis Tracking," in *Signal and Data Processing of Small Targets* SPIE vol. 5913, 2005.
- [5] Bar-Shalom, Y., and Campo, L., "The Effect of the Common Process Noise on the Two-Sensor Fused-Track Covariance," correspondence in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, pp. 803–805, 1986.
- [6] Bar-Shalom, Y., and Huimin, C., "Removal of Out-of-Sequence Measurements from Tracks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, pp. 612–619, 2009.
- [7] Bar-Shalom, Y., and Li, X.R., *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, 1995.
- [8] Blackman, S., and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 1999.
- [9] Blackman, S., "Multiple Hypothesis Tracking for Multiple Target Tracking," *IEEE A&E Systems Magazine*, vol., 19, no. 1, 2004.
- [10] Blair, W.D., "Radar Tracking Algorithms," Chapter 19 in *Principles of Modern Radar: Basic Principles*, Scitech, 2010.

- [11] Blair, W.D., *Monopulse Processing for Tracking Unresolved Targets*, Ph.D. dissertation, Dahlgren, VA, 1997.
- [12] Castanon, D., "New Assignment Algorithms for Data Association," in *Signal and Data Processing of Small Targets* SPIE vol. 1698, 1992.
- [13] Daum, F., "A Cramer-Rao Bound for Multiple Target Tracking," in *Signal and Data Processing of Small Targets* SPIE vol. 1481, 1991.
- [14] Ehrman, L., and Blair, W.D., "Using Target RCS When Tracking Multiple Rayleigh Targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, pp. 701–716, 2010.
- [15] Ehrman, L. and Blair, W.D., "Impact of Non-Coherent Pulse Integration on RCS-Assisted Tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, pp. 1573–1579, 2009.
- [16] Ehrman, L., Burns, P., and Blair, W.D., "Determining the Optimal Time Frame for Multisensor Track Correlation," in *Proceedings of Signal Processing of Small Targets*, SPIE vol. 6969, 2008.
- [17] Ehrman, L., Register, A.H., and Blair, W.D., "Measurement-to-Track Cost Assignment Computation When Tracking with an IMM Estimator," in *IEEE Aerospace Conference Proceedings*, March 2005.
- [18] Giannitrapani, A., Ceccarelli, N., Scortecchi, F., and Garulli, A., "Comparison of EKF and UKF for Spacecraft Localization via Angle Measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, pp. 75–84, 2011.
- [19] Jonker, R., and Volgenant, A., "A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems," *Computing*, vol. 38, pp. 325–340, 1987.
- [20] Kalman, R., "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME Journal of Basic Engineering*, 1960.
- [21] Kamen, E.W., "Multiple Target Tracking Based on Symmetric Measurement Equations," *IEEE Transactions on Automatic Control*, vol. 37, pp. 371–374, 1992.
- [22] Kirubakaran, T., and Bar-Shalom, Y., "Low Observable Target Motion Analysis Using Amplitude Information," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, pp. 1367–1384, 1996.
- [23] Kurien, T., and Liggins, M., "Report-to-Target Assignment in Multisensor Multitarget Tracking," in *Proceedings of the 27th IEEE Conference on Decision and Control*, pp. 2484–2488, 1988.
- [24] Kurien, T., and Washburn, R., "Multiobject Tracking using Passive Sensors," in *Proceedings of the American Control Conference*, pp. 1032–1038, 1985.
- [25] Lerro, D., and Bar-Shalom, Y., "Interacting Multiple model Tracking with Target Amplitude Feature," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, pp. 494–509, 1993.
- [26] Li, X.R., and Bar-Shalom, Y., "Multiple-Model Estimation with Variable Structure," *IEEE Transactions on Automatic Control*, vol. 41, pp. 478–493, April, 1996.
- [27] Li, X.R., and Bar-Shalom, Y., "A Hybrid Conditional Averaging Technique for Performance Prediction of Algorithms with Continuous and Discrete Uncertainties," in *Proceedings of the American Control Conference*, pp. 1530–1534, 1994.
- [28] Li, X.R., and Bar-Shalom, Y., "Performance Prediction of the Interacting Multiple Model Algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 755–771, 1993.

- [29] Mahler, R., “PHD Filters of Higher Order in Target Number,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, pp. 1523–1543, 2007.
- [30] Mallick, M., and La Scala, B., “IMM for Multisensor Ground Target Tracking with Variable Measurement Sampling Intervals,” in *Proceedings of the 2006 International Conference on Information Fusion*, 2006.
- [31] Mazor, E., Averbuch, A., Bar-Shalom, Y., and Dayan, J., “Interacting Multiple Model Methods in Target Tracking: A Survey,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, pp. 103–123, 1998.
- [32] Moon, T., and Stirling, W., “The Kalman Filter,” Chapter 13 in *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, 2000.
- [33] Mori, S., et. al., “Tracking Performance Evaluation-Prediction of Track Purity,” in *Signal and Data Processing of Small Targets* SPIE vol. 1089, 1989.
- [34] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 2002.
- [35] Poore, A., and Yan, X., “Use of K-Near Optimal Solutions to Improve Data Association in Multi-Frame Processing,” in *Proceedings of SPIE*, vol. 3809, 1999.
- [36] Poore, A., Robertson, A.J., and Shea, P., “A New Class of Lagrangian-Relaxation-Based Algorithms for Fast Data Association in Multiple Hypothesis Tracking Applications,” in *Proceedings of SPIE*, vol., 2484, 1995.
- [37] Poore, A., Rijavec, N., Liggins, M., and Vannicola, V., “Data Association Problems Posed as Multidimensional Assignment Problems: Problem Formulation,” in *Proceedings of SPIE*, vol. 1954, 1993.
- [38] Reid, D., “An Algorithm for Tracking Multiple Targets,” *IEEE Transactions on Automatic Control*, vol. 24, pp. 843–854, 1979.
- [39] Ristic, B., Arulampalam, S., and Gordon, N., *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, 2004.
- [40] Rothrock, R., and Drummond, O., “Performance Metrics for Multiple-Sensor Multiple-Target Tracking,” in *Signal and Data Processing of Small Targets* SPIE vol. 4048, 1999.
- [41] Ruan, Y., and Hong, L., “Feature-Aided Tracking with GMTI and HRR Measurements via Mixture Density Estimation,” *IEE Proceedings on Control Theory and Applications*, vol. 153, pp. 342–356.
- [42] Sittler, R., “An Optimal Data Association Problem in Surveillance Theory,” *IEEE Transactions on Military Electronics*, vol. 8, pp. 125–139, 1964.
- [43] Waxman, M., and Drummond, O., “A Bibliography of Cluster (Group) Tracking,” in *Signal and Data Processing of Small Targets* SPIE vol. 5428, 2004.

15.7 | PROBLEMS

1. Plot the covariance ellipses for the following (notional) 2-D covariance matrices. Do the off-diagonal terms affect the results? Why or why not?

a.
$$\begin{bmatrix} 9 & 0 \\ 0 & 16 \end{bmatrix}$$

b.
$$\begin{bmatrix} 9 & 10 \\ 12 & 16 \end{bmatrix}$$

2. Common metrics for evaluating the size of a covariance are the square root of the covariance's trace (e.g., $m_1 = \sqrt{\text{Trace}(P)}$) and the volume of the covariance ellipse (e.g., $m_2 = \pi \sqrt{\prod_i \lambda_i}$, in 2-D).
 - a. For the two 2-D covariance matrices in problem 1, compute m_1 and m_2 .
 - b. Why is m_1 the same for both examples while m_2 varies?
3. What percentage of the time is the target contained within an accurate (i.e., statistically consistent) Kalman Filter track covariance in a 1-D problem? How does this change if tracking in three dimensions? How many sigmas are needed in 3-D to reach 97% containment?
4. Compute the residual (i.e., innovations) and its covariance in each of the following examples, using (15.11) and (15.12). For the sake of simplicity, assume that the measurements are generated by a fictitious sensor that magically measures position in Cartesian coordinates. Hence, $h(x)$ simply strips the position terms out of the state (x), which includes position and velocity. In which example would the measurement most influence the track update? Why?

$$\text{a. } x = \begin{bmatrix} 10 \text{ m} \\ 10 \text{ m} \\ 10 \text{ m} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \end{bmatrix}, P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2 \end{bmatrix}, z = \begin{bmatrix} 10.5 \text{ m} \\ 9.5 \text{ m} \\ 11.0 \text{ m} \end{bmatrix}, R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{b. } x = \begin{bmatrix} 10 \text{ m} \\ 10 \text{ m} \\ 10 \text{ m} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \end{bmatrix}, P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2 \end{bmatrix}, z = \begin{bmatrix} 10.5 \text{ m} \\ 9.5 \text{ m} \\ 11.0 \text{ m} \end{bmatrix}, R = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

$$\text{c. } x = \begin{bmatrix} 10 \text{ m} \\ 10 \text{ m} \\ 10 \text{ m} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \end{bmatrix}, P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2 \end{bmatrix}, z = \begin{bmatrix} 10.5 \text{ m} \\ 9.5 \text{ m} \\ 11.0 \text{ m} \end{bmatrix},$$

$$R = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}$$

5. Given the following measurement-to-track cost matrix, which measurement(s) would not gate with Track 2? What is the optimal association answer for this cost matrix?

	Measurement 1	Measurement 2	Measurement 3	Ø
Track 1	5	10	15	20
Track 2	10	30	12	20
Ø	20	20	20	∞

6. Given the following measurement-to-track cost matrix, which measurement(s) would not gate with Track 2? What is the optimal association answer for this cost matrix? How many tracks are coasted? How many measurements become candidates for new track initiation?

	Measurement 1	Measurement 2	Measurement 3	Ø
Track 1	5	10	15	20
Track 2	25	30	35	20
Ø	20	20	20	∞

7. Given two Swerling II targets with measurement SNRs of 15 dB and 21.2 dB, what is probability of error in measurement-to-track assignment using only the SNR feature, neglecting any pulse integration? How does this change if eight pulses are integrated?
8. Suppose that you are asked to design a tracker for an air-traffic-control system at a small airport. The aircraft may be observed while flying straight-and-level (when far away from the airport) or while maneuvering to land or take off. Aircraft arrive and depart the airport infrequently, making it extremely unlikely that more than one will ever be in track at the same time. If you have to choose one or the other because of cost constraints, would you design an IMM or a MHT for this application? Explain your rationale.
9. Give four reasons for sensor fusion.
10. Using the naïve equations for state and covariance track fusion (dubbed ‘naïve’ because they assume zero cross-correlation between the tracks), find the fused state (x_f) and covariance (P_f) in the following two examples.

$$P_f = \left(\sum_i P_i^{-1} \right)^{-1}, \quad x_f = P_f \left(\sum_i P_i^{-1} x_i \right)$$

How does the fused covariance relate to that of the contributing tracks?

$$\text{a. } x_1 = \begin{bmatrix} 10 \text{ m} \\ 10 \text{ m} \\ 10 \text{ m} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \end{bmatrix}, \quad P_1 = \begin{bmatrix} 2.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 12 \text{ m} \\ 12 \text{ m} \\ 12 \text{ m} \\ 1.75 \text{ m/s} \\ 1.75 \text{ m/s} \\ 1.75 \text{ m/s} \end{bmatrix},$$

$$P_2 = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.75 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.75 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.75 \end{bmatrix}, \text{ unbiased state estimates}$$

$$\text{b. } x_1 = \begin{bmatrix} 10 \text{ m} \\ 10 \text{ m} \\ 10 \text{ m} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \\ 2 \text{ m/s} \end{bmatrix}, \quad P_1 = \begin{bmatrix} 2.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 12 \text{ m} \\ 12 \text{ m} \\ 12 \text{ m} \\ 1.75 \text{ m/s} \\ 1.75 \text{ m/s} \\ 1.75 \text{ m/s} \end{bmatrix},$$

$$P_2 = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.75 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.75 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.75 \end{bmatrix},$$

$$\text{Sensor 2 has an uncompensated bias of } b_2 = \begin{bmatrix} 0.5 \text{ m} \\ 0.5 \text{ m} \\ 0.5 \text{ m} \\ 0 \text{ m/s} \\ 0 \text{ m/s} \\ 0 \text{ m/s} \end{bmatrix}.$$

PART V

Emerging Techniques

**CHAPTER 16 Human Detection With Radar:
Dismount Detection**

**CHAPTER 17 Advanced Processing Methods
for Passive Bistatic Radar System**

Human Detection With Radar: Dismount Detection

Sevgi Zübeyde Gürbüç

Chapter Outline

16.1	Introduction	705
16.2	Characterizing the Human Radar Return	710
16.3	Spectrogram Analysis of Human Returns	719
16.4	Technical Challenges in Human Detection	722
16.5	Exploiting Knowledge for Detection and Classification	727
16.6	Summary	729
16.7	Further Reading	729
16.8	References	730
16.9	Problems	736

16.1 | INTRODUCTION

Human detection, also known as pedestrian or dismount detection, has been accomplished with a variety of devices, including visual [1–5], acoustic [6,7], vibration/seismic [8–11], infrared [12–14], and laser [15] sensors in addition to radar. Many of these systems apply sensor fusion techniques to combine the strengths of sensor modalities in a complementary fashion. Examples include radar in conjunction with infrared sensors [14] or video cameras [16]. Although the augmentation of radar with other sensors may be useful in some applications, the development of autonomous human detection capabilities for radar is critical in many applications, especially those involving security and defense.

Radar offers several unique advantages over other sensors, namely, the ability to operate remotely, far away from environments dangerous to operators, and in virtually all weather and lighting conditions. Through-the-wall radar can even provide the surveillance of indoor environments from the outside of a structure. In contrast, visual, acoustic, infrared, and seismic sensors all require being placed in proximity to the target, without any obstacles impeding the line of sight. Visual sensors can operate only during the daytime. Thus, radar remains the principal sensor of choice for many surveillance applications involving human detection.

Over the years several platforms capable of performing human detection, in addition to detection of vehicles, have been developed. One of the older systems still in use by the U.S. Army is the AN/PPS-5 Ground Surveillance Radar, originally developed in 1966 by the Eaton Corporation. It is a portable system operating in a band of 8.8–9 GHz, capable

TABLE 16-1 ■ Performance Specifications of Currently Available Ground Surveillance Radar Platforms

System	Band	Maximum Range	Range Accuracy	Minimum Detectable Velocity
AN/PPS-5	8.8–9 GHz	Personnel: 5 km Vehicles: 10 km	20 m	1.6 km/h (0.45 m/s)
AN/PPS-5C	10–20 GHz	Single Man: 7 km Small Vehicle: 16 km Large Vehicle: 23 km	8.3 m	1.25 m/s (radial)
SR Hawk	16.21–16.50 GHz	Single Man: 10 km Avg. Vehicle: 30 km	10 m	0.13–0.25 m/s (radial)
STS-700	26.5–40 GHz	Personnel Walking: 700 m Personnel Running: 700 m Personnel Low Crawl: 500 m Personnel Crawling: 700 m Personnel Swimming: 500 m Vehicles: 700 m	Not Given	0.1 m/s
STS-1400	26.5–40 GHz	Personnel Walking: 1400 m Personnel Running: 1400 m Personnel Low Crawl: 500 m Personnel Crawling: 1400 m Personnel Swimming: 500 m Vehicles: 1400 m	Not Given	0.1 m/s
STS-2800	26.5–40 GHz	Personnel Walking: 1900 m Personnel Running: 1900 m Personnel Low Crawl: 500 m Personnel Crawling: 1900 m Personnel Swimming: 500 m Vehicles: 2800 m	Not Given	0.1 m/s
STS-4400	Not Given	4400 m (person or vehicle)	Personnel: 0.4 m 30 mph Vehicle: 1.5 m	0 m/s
STS-12000	Not Given	Personnel: 9.7 km Vehicles: 12 km	3 m	0 m/s

Sources: [17–23].

of detecting personnel at ranges up to 5 km and vehicles up to 10 km. This radar has undergone several upgrades; the most recent (AN/PPS-5D) extends personnel detection to 10 km and vehicle detection to 20 km [17–19]. A next-generation ground surveillance radar developed by SRC, Inc., the SR Hawk, operates at a band of 16.21–16.50 GHz with a maximum detection range of 10 km for a single walking person and 30 km for a typical vehicle [20]. The STS family of radars developed by ICX Technologies also offers similar performance, with the STS-12000 Long-Range Perimeter Surveillance Radar offering a maximum range of 9.7 km for personnel and 12 km for vehicle detection [21–23]. Key performance parameters of currently available ground surveillance radars are summarized in Table 16-1.

These figures, however, represent the best possible performance of the radar, often under near ideal conditions. In typical situations, human targets are easily masked by ground clutter. This ground clutter is the result of strong reflections from the surrounding

environment that may mask the generally weaker return from a person. Since people are typically slow-moving targets, humans also often fall below the minimum detectable velocity (MDV) of ground moving target indication (GMTI) radars. Moreover, these ground surveillance radars also offer no information beyond that of a target being detected at a certain range, azimuth angle, and radial velocity. These radars are usually used in tandem with video cameras that can zoom in on the detected object and thereby determine what it is, what it is doing, and whether it poses a threat. Thus, in addition to the development of more sophisticated, higher performance human detection algorithms, which function in even the most adverse conditions, the associated automatic identification, activity classification, and tracking algorithms remain equally important research topics.

With these goals in mind, there are several promising programs under way to develop aerial dismount detection radar. The upgrade to the Lynx Advanced Multi-Channel Radar (AMR) [24], currently under development by General Atomics Aeronautical Systems in cooperation with Sandia National Laboratories, was successfully tested for dismount detection during flights conducted in May 2010 and achieved operational availability by the end of 2010 [25]. The Lynx system is a synthetic aperture radar (SAR) with GMTI capability that exploits space-time adaptive processing (STAP) for dismount detection. Lynx, which is designed for the Predator and Predator B type unmanned aerial systems (UASs), is being evaluated as a possible upgrade to the Army's MQ-1C Sky Warrior/Gray Eagle unmanned aircraft system, which is equipped with an Northrop-Grumman AN/ZPY-1 STARlite small tactical radar [26, 27].

Another promising program is the Vehicle and Dismount Exploitation Radar (VADER) [28, 29] being designed by Northrop Grumman, which completed its first round of testing in 2009 [30]. VADER has three operating modes: SAR, GMTI, and dismount moving target indication. The specific goal of VADER is to "track moving dismounted individuals and characterize suspicious actions" from an altitude of about 25,000 ft (7.62 km), such as people planting improvised explosive devices (IEDs) [31]. Additional features envisioned for VADER include long-duration vehicle tracking, SAR coherent change detection (CCD), and motion pattern analysis. The VADER system, which is physically encased in a pod about 1.5 meters long, was designed to fit beneath the wing of a Gray Eagle medium-altitude UAS but is also being considered for mounting on YMQ-18A A160T Hummingbird UASs.

Two other programs worth attention are the Affordable Adaptive Conformal ESA Radar (AACER) program and the Foliage Penetration, Reconnaissance, Surveillance, Tracking and Engagement Radar (FORESTER) program. AACER was awarded to Raytheon's Space and Aerospace Systems (SAS) for the development of a multifunctional, multifrequency system capable of ground moving target detection and tracking. In addition to dismount detection, the radar also includes SAR imaging, and high data-rate Ka band communications [32]. The FORESTER program was awarded to Syracuse Research Corporation to accomplish dismount and vehicle detection even in heavily covered areas, such as forests [33]. The FORESTER program has been tested on the EH-60 Black Hawk, and the A160T Hummingbird UAS in Fort Stewart, GA in 2009 and over Belize in 2010 [34].

Thus, human-dismount detection is an integral capability of many of the radar systems currently under development. This chapter is intended to examine in detail the modeling and signal processing techniques that have made these recent advances possible. In Section 16.2, the salient characteristics of human motion, kinematic models based on gait analysis, and expected radar return will be discussed. With this foundation, spectrogram-based

human identification techniques will be presented in Section 16.3. The technical challenges involved with human detection will be quantitatively examined in Section 16.4. Section 16.5 outlines some of the recent novel ideas put forward to advance dismount detection technologies, while Section 16.6 provides a brief summary of the chapter, Section 16.7 recommends sources for additional reading, and Section 16.8 provides a detailed list of relevant references. A number of exercises designed to test the reader's comprehension are included in Section 16.9.

16.1.1 Key Points

- Human targets are difficult to detect due to their low radar cross section and low velocity.
- Humans often fall below the MDV of ground moving target indicators and are easily masked by ground clutter.
- Models of human kinematics, such as the Boulic-Thalmann walking model, are being used to analytically compute the expected human radar return.
- The development of human models including a variety of human motion, such as running, jumping, or crawling, as well as accurate human cross section models are still a subject of significant research.
- The unique nature of human bipedal motion is the cause of micro-Doppler features in the human spectrogram, which may be used to identify and classify detected targets.
- STAP is a key technique for mitigating the effects of clutter but by itself is not sufficient to ensure the detection of dismounts in adverse clutter environments.
- The nonlinear phase history of human targets, caused by micro-Doppler effects, results in an inherent signal-to-noise ratio (SNR) loss in the radar detector. The consequence is a loss in dismount detection performance.
- Novel techniques recently proposed include exploiting results from gait analysis as a priori knowledge that can be incorporated in human detection and classification algorithms.

16.1.2 Notation

A = sinusoidal approximation to human model parameter: factor in amplitude term

A_L = Boulic-Thalmann model parameter: amplitude of lateral translation of torso

A_{FB} = Boulic-Thalmann model parameter: amplitude of forward/backward translation of torso

ATR = appendage to torso ratio

a_t = received signal amplitude

C_1 = sinusoidal approximation to human model parameter: factor relating to target range

C_2 = sinusoidal approximation to human model parameter: amplitude of torso motion

C_3 = sinusoidal approximation to human model parameter: torso frequency

C_4 = sinusoidal approximation to human model parameter: torso phase

c = speed of light

d_{bp} = dimension of a human body part

D_b = Boulic-Thalmann model parameter: duration of balance

D_c = Boulic-Thalmann model parameter: duration of cycle

D_{ds} = Boulic-Thalmann model parameter: duration of double support

D_s = Boulic-Thalmann model parameter: duration of support

$D_s\%$ = Boulic-Thalmann model parameter: relative duration of support
 f_c = transmitted center frequency
 G = antenna gain
 γ = chirp slope
 H_c = height of a cylindrical target
 H_e = height of an ellipsoidal target
 HT = Boulic-Thalmann model parameter: height of thigh
 \mathbf{h} = human motion vector
 K = human model parameter: number of parts into which human body is divided
 λ = signal wavelength
 M = sinusoidal approximation to human model parameter: slope of linear phase component
 μ_{bp} = mean value of Gaussian-distributed body part dimension
 N = total number of pulses transmitted
 n = pulse number
 OS_L = Boulic-Thalmann model variable: lateral translation of torso
 OS_V = Boulic-Thalmann model variable: vertical translation of torso
 OS_{FB} = Boulic-Thalmann model variable: forward/backward translation of torso
 ϕ_{FB} = Boulic-Thalmann model parameter: phase shift of forward/backward translation of torso
 P_t = transmit power
 Δr = distance between the center of an ellipsoid and the point on the surface of the ellipsoid at which an incident wave first makes contact
 R = radar slant range
 R_c = radius of base of a cylindrical target
 R_e = radius of an ellipsoidal target for two equivalent sides
 RLC = Boulic-Thalmann model parameter: relative length of walking cycle
 R_s = radius of a spherical target
 RV = Boulic-Thalmann model parameter: average walking velocity
 \mathbf{r}_1 = initial target vector
 \mathbf{r}_N = vector pointing to target location after N pulses transmitted
 r_b = range bin of peak pulse compressed human return
 s_h = received radar signal for a human target
 s_r = received radar signal for a point target
 σ = target radar cross section
 σ_{bp} = variance of Gaussian-distributed body part dimension
 σ_c = radar cross section for a cylinder
 σ_e = radar cross section for an ellipsoid
 σ_s = radar cross section for a sphere
 t = total time elapsed since transmission of first pulse
 \hat{t} = time elapsed within a single pulse repetition interval
 $t\%$ = Boulic-Thalmann model parameter: time relative to start of walking cycle
 T = pulse repetition interval
 τ = pulse width
 t_d = round-trip time delay of signal between antenna and target
 θ_e = angle with respect to the height axis for a ellipsoidal target
 x_p = pulse compressed human return
 $x_{p,sin}$ = sinusoidal approximation to pulse compressed human return

16.1.3 Acronyms

ATR	appendage-to-toe ratio
CCD	coherent change detection
CPI	coherent processing interval
FFT	fast Fourier transform
GMTI	ground moving target indication
HS	heel strike
IED	improvised explosive devices
MDV	minimum detectable velocity
PDF	probability density function
PRI	pulse repetition interval
RCS	radar cross section
SAR	synthetic aperture radar
SIR	signal-to-interference ratio
SNR	signal-to-noise ratio
STAP	space-time adaptive processing
TL	toe lift
UAS	unmanned aerial system

16.2 | CHARACTERIZING THE HUMAN RADAR RETURN

Humans are complicated targets due to the intricate motion of many parts, all moving at different speeds and along different trajectories, depending on the type of activity engaged in, such as walking, running, jumping or playing sports [35–38]. Research has shown that human kinematics even change depending on whether or not a person is carrying a load, and how this load is being carried [39–41]. Indeed, some optimistic researchers have even stated that the periodic, bipedal nature of human walking is so unique that in the future human gait may be used as a biometric parameter for identification, similar to the way fingerprints and hand geometry are used today [42–44]. While the wide variation in human activity makes the development of a universal model characterizing human motion difficult, it also provides a basis for the development of radar signal processing algorithms capable of discriminating a variety of human activities.

When a radar signal interacts with a target, the reflected signal depends on a number of factors, including the target's RCS and time-varying range, among others. For rigid-body targets such as the chassis of a vehicle, virtually all points comprising the target move at the same speed relative to the radar, leading to a constant Doppler shift in the frequency of the reflected signal received by the radar. If, however, there is motion, vibration, or rotation, in addition to bulk translation, then sidebands about the target's bulk Doppler frequency are generated [45–47]. Such micro-Doppler signals may be the result of reflections from the wheels of vehicles, aircraft propellers, and helicopter blades, as well as human motion.

Early efforts to characterize the human micro-Doppler signature began in 1997 with work by Weir [48,49], in which the velocity profile of a human target (what Weir termed a gait velocitygram) was observed to contain distinct features that could be used to garner information about the gait. Later work by Geisheimer [50] involved the use of chirplet

transforms to characterize the gait signature; however, since then, most researchers have used spectrogram analysis as a means of visualizing micro-Doppler.

The modeling, simulation, and understanding human spectrograms is important in the development of dismount detection algorithms for radar. Model-based approaches [51–53] exploit mathematical models of human kinematics to analytically derive the spectrogram of human radar returns. Alternatively, computer models derived from virtual reality animation data have been successfully used to simulate human micro-Doppler [54–57]. One widely-used 3-D human motion data library is that compiled by the Carnegie Mellon University (CMU) Motion Research Laboratory, which is available online at [58]. The CMU library is based on data captured from video camera coverage of real human motion. Using this data, a 3-D motion model for a human body comprised of 30 bones is created and animated. Thus, the motion is not modeled or approximated, but instead directly measured and stored. Motion-capture based models have been successfully used in simulations that enable the assessment of ground reflections and multipath [59], multiple interactions between different body parts, as well as through-the-wall propagation [60] on human micro-Doppler signatures. In comparison, model-based approaches are limited in modeling such effects; however, they are computationally fast and reasonably accurate, while also providing insight into the underlying physical process that makes human spectrograms unique. In the remainder of this section, a mathematical formulation of the expected human radar return is derived. This model will be used later in Section 16.3 to derive and analyze the spectrograms of human targets.

16.2.1 Expected Radar Return from a Human Target

Consider a radar antenna transmitting a series of chirped pulses at constant intervals in time and space while moving along a straight path. In general, the received radar signal is a time-delayed version of the transmitted chirp signal (Figure 16-1). The received radar signal for a point target may be expressed as

$$s_r(n, t) = a_t \text{rect}\left(\frac{\hat{t} - t_d}{\tau}\right) e^{j[-2\pi f_c t_d + \pi \gamma (\hat{t} - t_d)^2]} \quad (16.1)$$

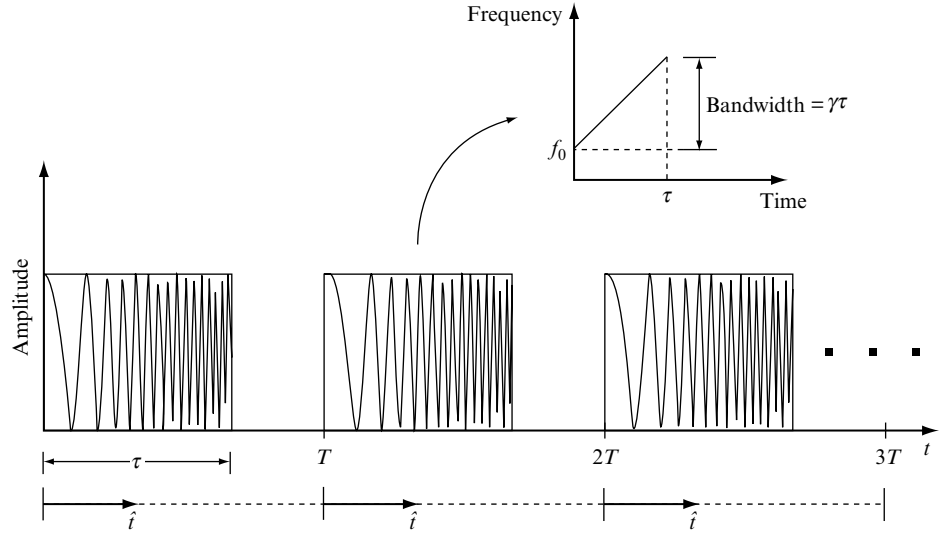
where n is the pulse number; t is the total time elapsed since transmission of the first pulse and may be defined in terms of the pulse repetition interval (PRI), T , and the time relative the start of each PRI, \hat{t} , as $t = T(n - 1) + \hat{t}$; τ is the pulse width; c is the speed of light; γ is the chirp slope; f_c is the transmitted center frequency; and t_d is the round-trip time delay between antenna and target, defined in terms of the target slant range, R , as $t_d = 2R/c$. The signal amplitude, a_t , is given by the radar range equation as

$$a_t = \frac{G\lambda\sqrt{P_t\sigma}}{(4\pi)^{1.5}R^2} \quad (16.2)$$

where G is the antenna gain, λ is the signal wavelength, P_t is the transmit power, σ is the target RCS, and R is the range from the radar to the target. Here, it is also assumed that there are no losses due to atmospheric attenuation or system noise.

Human targets, of course, are not point targets but may be viewed as the collection of a large number of point targets residing upon the surface of the human body. An important result is that the spectrogram from the entire human body can be well approximated by the sum of the spectrograms from the constituent body parts [61], which implies that for

FIGURE 16-1 ■
Illustration of
transmitted pulsed
Doppler chirp signal
and key parameter
definitions.



human targets the principle of superposition is valid. Moreover, work by Van Dorp [51] has shown that by (1) dividing the human body into just 12 parts, (2) modeling each part as a point target, and (3) summing the responses, a mathematical model for the human response that matches remarkably well with measured spectrogram data can be produced. Thus, a mathematical approximation to the human radar return may be written from (16.1) as

$$s_h(n, t) = \sum_{i=1}^K a_{t,i} \text{rect}\left(\frac{\hat{t} - t_{d,i}}{\tau}\right) e^{j[-2\pi f_c t_{d,i} + \pi \gamma (\hat{t} - t_{d,i})^2]} \quad (16.3)$$

where K is the total number of parts into which the body is divided, and i is an index referring to each body part ($1 \leq i \leq K$). Embedded within this equation are two factors that are directly impacted by human modeling: the time-varying range R_i of each body part to the radar (in $t_{d,i} = 2R_i/c$), and RCS of each component (i.e., the factor σ in $a_{t,i}$).

16.2.2 Human Kinematic Models

Researchers have pursued a variety of approaches in modeling human kinematics, most of which revolve around decomposing the human body into a finite number of parts and computing the time-varying range for each part using kinematic models. An alternative approach has been taken by R. G. Raj, who has pointed out that in many cases human activity occurs in real-time unconstrained environments, where the type of activity or model parameters are not known a priori, leading to a mismatch between the true activity and motion model used. Thus, Raj proposed directly modeling and classifying the time-frequency motion curves using a Gaussian *g-snake* model [62, 63].

Nevertheless, kinematic models have provided valuable insight and utility in the detection and identification of dismounts and remain the most widely explored approach today. For example, He [64] recently proposed a modified version of the linear-rigid model developed by Zhang [65] in which the head and torso were modeled as a single point scatterer, the arms as independent cylinders, and the legs as two interconnected cylinders. The periodic arm motion was modeled as being that of a pendulum, while the leg motion was modeled by a combination of vibrations and rotations.

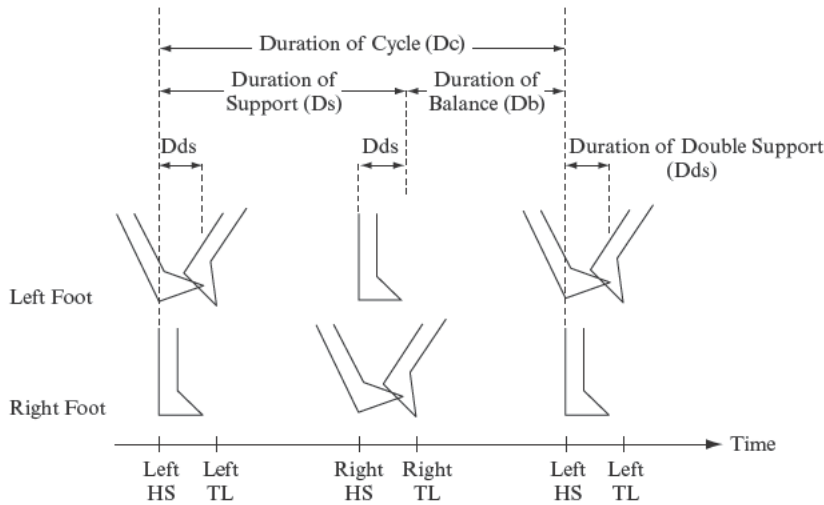


FIGURE 16-2 ■
Temporal structure
of a typical human
walking cycle.

The model that has garnered the most attention and use, however, is the Boulic-Thalman model [66], first proposed in 1990 and successfully used in human spectrogram modeling by Van Dorp [51] in 2003. The Boulic-Thalman model is a mathematical parametrization of human walking based on biomechanical experimental data. As shown in Figure 16-2, the walking cycle is temporally modeled as being made up of two periods of support and two periods of balance. The cycle begins with a heel strike (HS) of the left foot and ends with the lifting of the toe (TL) of the right foot. The duration of support is defined as the time during which a leg supports the body's weight, while the duration of balance encompasses the period during which a leg is raised in the air. Thus, the time between HS and TL is a period of double support.

These temporal periods are mathematically modeled by Boulic and Thalman as follows. The duration of cycle, D_c , in seconds is given by

$$D_c \equiv \frac{RLc}{RV} \quad (16.4)$$

where RLc is the relative length of the walking cycle, and RV is the average walking velocity, $v(m/s)$, normalized by the thigh height, $HT(m)$, as shown in Figure 16-3. Thus,

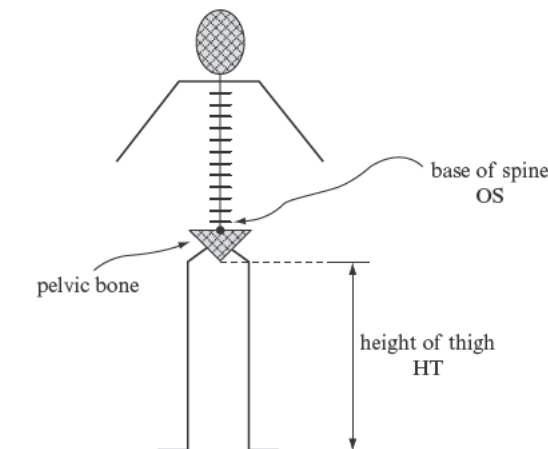


FIGURE 16-3 ■
Definition of body
coordinate system.

$RV = v/HT$. The parameter RLc can be expressed in terms of RV as $RLc = 1.346\sqrt{RV}$, leading to the following simplification in the expression for Dc :

$$Dc = \frac{1.346}{\sqrt{RV}} \quad (s) \quad (16.5)$$

The duration of support, Ds , the duration of double support, Dds , and duration of balance are all linearly dependent upon the cycle duration and are given as

$$Ds = 0.752Dc - 0.143 \quad (s) \quad (16.6)$$

$$Dds = 0.252Dc - 0.143 \quad (s) \quad (16.7)$$

$$Db = 0.248Dc + 0.143 \quad (s) \quad (16.8)$$

A relative time parameter, $t\%$, is defined as

$$t\% = \frac{t}{Dc} \quad (16.9)$$

while the relative duration of support, $Ds\%$, is defined as

$$Ds\% = \frac{Ds}{Dc} \quad (16.10)$$

In both (16.9) and (16.10), the term “relative” refers to the time elapsed as being a percentage of the overall walking cycle duration.

These temporal values are used in the parametric model, which produces the 3-D kinematic equations for the time-varying change in position of certain points on the body as well as the time-varying angular motion of elbow, knee, and ankle joints. All motion is described relative to the origin, defined at the bottom of the spine, OS (as shown in Figure 16-3). Note that the thigh height, which is a critical parameter in the model, is defined as the distance between the ground and the top of the thigh bones, while OS is located higher up at the base of the spine. The kinematics governing pelvic rotations and flexing at the hip are thus also modeled.

Consider the equations given by the Boulic-Thalman model for the periodic trajectories that govern the motion of the origin, OS , for the duration of one walking cycle:

Vertical translation:

$$OS_V = 0.015RV (\sin 2\pi (2t\% - 0.35) - 1) \quad (16.11)$$

Lateral translation:

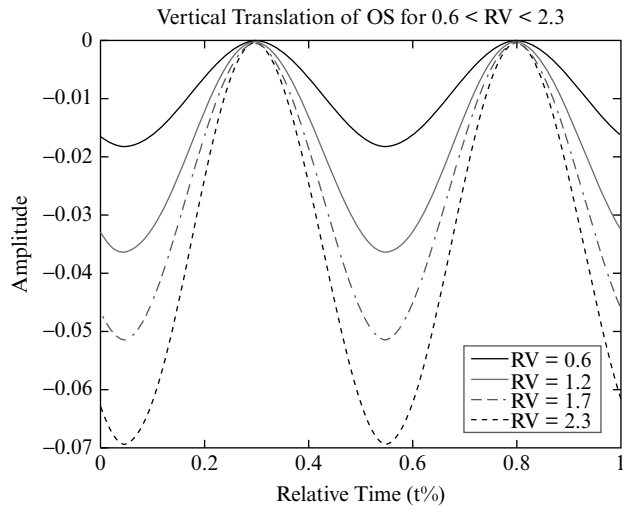
$$OS_L = A_L \sin 2\pi (t\% - 0.1), \text{ where } A_L = \begin{cases} -0.032 & \text{for } RV > 0.5 \\ -0.128RV^2 + 0.128RV & \text{for } RV < 0.5 \end{cases} \quad (16.12)$$

Forward/backward translation:

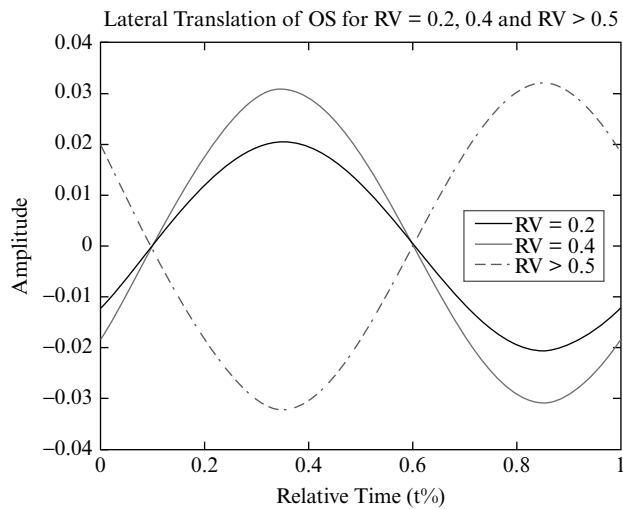
$$OS_{FB} = A_{FB} \sin 2\pi (2t\% - 2\phi_{FB}) \quad (16.13)$$

$$\text{where } \phi_{FB} = 0.625 - Ds\% \text{ and } A_{FB} = \begin{cases} -0.021 & \text{for } RV > 0.5 \\ -0.084RV^2 + 0.084RV & \text{for } RV < 0.5 \end{cases}$$

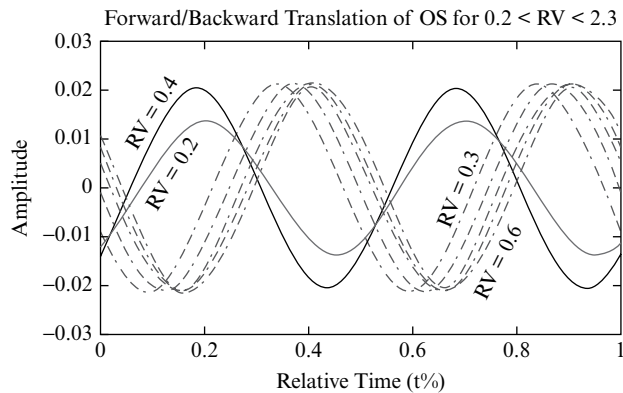
A plot of the trajectories of OS for varying values of RV is shown in Figure 16-4. These curves are significant because when these components are combined to form the overall



(a)



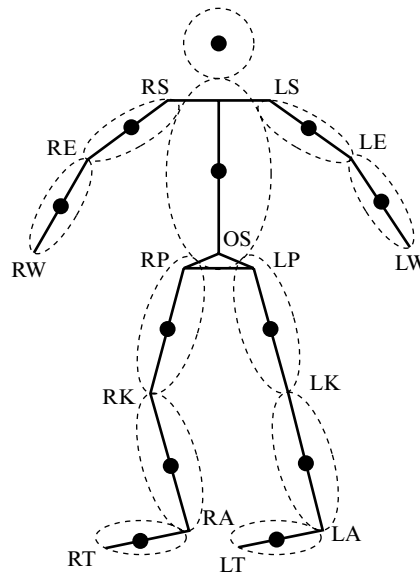
(b)



(c)

FIGURE 16-4 ■
Periodic trajectories
for OS. (a) Vertical
translation.
(b) Lateral
translation.
(c) Forward/
backward translation

FIGURE 16-5 ■
12-point target
representation of
human body. (From
Gürbüz et al. [69].
With permission.)



motion vector for *OS*, the combined trajectory overlaps with the torso response component of the human spectrogram. The trajectories for other body parts, such as the hands and feet, are also computable from the Boulic-Thalman model and similarly coincide with the components observed in the human spectrogram. Thus, kinematics provides a physical explanation for the unique shape of human spectrograms.

Unfortunately, most of the human body kinematics is not described in terms of neat and tidy equations, such as those describing the kinematics of *OS*, but in terms of relatively complex graphs. Thus, the Boulic-Thalman walking model is not a closed-form model. The equations and graphs provided in [66] can be easily used to animate a multiple point-target type human stick figure, and thereby generate a simulated response of a human target to radar; however, its closed-form nature limits its utility in certain situations, such as mathematically deriving the radar detection performance for human targets. Nevertheless, the Boulic-Thalman model describes human walking sufficiently well that it may be used to classify detected dismounts [67, 68] and to simulate the spectrograms of the expected human target response [69, 70], as will be seen later in this section.

16.2.3 Computing Time-Varying Ranges with the Boulic-Thalman Model

Consider the 12-point target representation of the human body shown in Figure 16-5. The head, upper arms, lower arms, torso, upper legs, lower legs, and feet are each modeled as a point target located at the centroid of the body part. The stick figure from the Boulic-Thalman model is superimposed upon the point target representation and used in conjunction with information on body sizes to compute the time varying range of each point target. The head and torso are considered to be rigidly attached to each other. Given the limb sizes, the angular variation at the joints may be converted into time-varying distances for the remaining point targets by elementary trigonometry.

A listing of human body part dimensions is provided in Table 16-2. Using interpolation techniques, the dimensions given in Table 16-2 may be scaled to achieve any sized human desired. Because overall human heights approximately vary with a Gaussian distribution,

TABLE 16-2 ■ Human Dimensions in Meters

(Meters)	95% Male	5% Male	95% Female	5% Female
Initial OS	1.180	1.005	1.085	0.930
Height of thigh	1.000	0.840	0.885	0.740
Torso height	0.505	0.390	0.460	0.360
Head height	0.205	0.180	0.190	0.165
Hip width	0.405	0.310	0.435	0.310
Hip height	0.295	0.195	0.280	0.185
Thigh length	0.550	0.440	0.530	0.435
Lower leg length	0.490	0.395	0.445	0.355
Foot length	0.285	0.240	0.255	0.215
Shoulder length	0.215	0.1825	0.1925	0.1625
Shoulder-elbow length	0.395	0.330	0.360	0.300
Elbow-wrist length	0.305	0.265	0.270	0.240
Torso diameter	0.357	0.290	0.340	0.2675
Upper arm diameter	0.095	0.080	0.085	0.070
Lower arm diameter	0.07125	0.060	0.06375	0.0525
Thigh diameter	0.185	0.135	0.180	0.125
Leg diameter	0.1387	0.1013	0.135	0.9375
Foot diameter	0.110	0.085	0.100	0.080

Source: From the RoyMech website [71] (with permission).

it follows that the body part dimensions also scale with a Gaussian distribution. In general, the k th percentile is defined as the threshold for which $k\%$ of the data has a value equal to or less than the threshold. Thus, in Table 16-2, the value of 0.840 meters for the height of thigh of a 5th percentile male means that 5% of male humans have a height less than or equal to 0.840 meters.

To scale the body part dimensions for any size human, the mean and variance of body part sizes, for males and females separately, must be computed. Define the dimension of any body part, d_{bp} , as

$$d_{bp} = \mu_{bp} + \alpha_k \sigma_{bp} \quad (16.14)$$

where μ_{bp} and σ_{bp} are the mean and variance of the distribution of that body part. The parameter α_k is the threshold for the k th percentile computed for a zero mean and unit variance standard normal distribution,

$$\frac{k}{100} = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (16.15)$$

a value that may be easily found in mathematical tables of the Gaussian cumulative distribution function.

Thus,

$$d_{bp,95\%ile} = \mu_{bp} + \alpha_{95} \sigma_{bp} \quad (16.16)$$

$$d_{bp,5\%ile} = \mu_{bp} + \alpha_5 \sigma_{bp} \quad (16.17)$$

a two equation two unknown system for which the body part mean and variance can be computed. Then, the body part dimension for any size human is found from Equation (16.14), this time substituting the value of α_k corresponding to the desired percentile.

As explained in detail in the next section, the body part dimensions computed from Table 16-2 are used not just to position the point scatterers modeling each body part, but also to compute approximations to the RCS of each body part.

16.2.4 Modeling Human Radar Cross Section

The other important human dependent factor in (16.3) is the modeling of the RCS of a human in motion. The RCS of an object depends on many factors, such as the shape of the object; the material composing the object; the type, frequency, and polarization of the radar; and angle of incidence. Human RCS is difficult to ascertain because a person's posture, relative orientation, incidence angle, and distance relative to the radar constantly change during any type of activity. Moreover, human size varies significantly, ranging from that of a small child to that of tall man. Even the type of material in the clothes a person is wearing can affect the reflectivity properties and, hence, observed RCS [72]. In general, however, human RCS is significantly lower than the RCS of other typical targets of interest, such as vehicles. At microwave frequencies, the RCS of a man is about 1 m^2 (0 dBsm), whereas the RCS of a pickup truck is 200 m^2 (23 dBsm), a factor of 200 greater [73].

The RCS modeling of humans is further complicated by the presence of clutter and multipath reflections from the ground and between body parts in motion. The scattering from a human target cannot be considering in isolation from such effects, so that essentially it is the *composite response* of the human together with the ground that must be modeled. Some recent works that investigate the impact of ground reflections and multipath on the human return include [59], [74], and [75].

Due to the lack of definitive studies leading to a comprehensive model for human RCS, many researchers have instead opted to approximate the complex human RCS with the practical, computationally simple Marcum (Swierling Case 0) model. This approach assumes a nonfluctuating (i.e., constant amplitude) model for human RCS and computes the scattering amplitude from the shape of the body part; for example, a spherical scatterer for the head and cylindrical or ellipsoidal scatterers for the torso and limbs [51, 69, 70]. More specifically, for a sphere with radius R_s (m) the complex scattering value is given by [76]

$$\sqrt{\sigma_s} = R_s \sqrt{\pi} \left[\left(1 + \frac{\lambda}{j4\pi R_s} \right) e^{-j\left(\frac{4\pi R_s}{\lambda}\right)} - \frac{\lambda}{j4\pi R_s} \right] \quad (16.18)$$

For a cylinder with radius R_c (m), height H_c (m), and broadside incidence angle θ_c (rad), the complex scattering value is given by [76]

$$\sqrt{\sigma_c} = -jH_c \sqrt{\cos(\theta_c) \frac{2\pi R_c}{\lambda}} \text{sinc} \left(\frac{2\pi H_c}{\lambda} \sin \theta_c \right) e^{j\frac{2\pi}{\lambda}(r+R_c \cos \theta_c)} e^{j\frac{\pi}{4}} \quad (16.19)$$

where r is the range to the target from the radar. The complex scattering value for an ellipsoid is approximately given by [51] as

$$\sqrt{\sigma_e} = \frac{R_e^2 \frac{1}{2} H_e \sqrt{\pi}}{(R_e \sin \theta_e)^2 + \left(\frac{1}{2} H_e \cos \theta_e \right)^2} e^{j\left(\frac{2\pi}{\lambda}\right) \Delta r} \quad (16.20)$$

where R_e is the radius of the ellipsoid for two equivalent sides, H_e is the height of the ellipsoid, θ_e is the angle with respect to the height axis, and Δr is the distance between the center of the ellipsoid and the point on the surface of the ellipsoid at which the incident wave makes first contact. More general expressions for the RCS of an ellipsoid can be found in [77, 78].

TABLE 16-3 ■ Summary of Swerling Target Fluctuation Model Characteristics

Swerling Model	Target Type	Fluctuation	PDF of RCS
I	Many independent scatterers of equal size	scan-to-scan	Rayleigh
II		pulse-to-pulse	
III	One large scatterer surrounded by many smaller scatterers	scan-to-scan	Chi-square, degree 4
IV		pulse-to-pulse	

However practical the nonfluctuating Marcum model for human RCS may be, results from a 2007 study [79] show that fluctuating models may better represent the true human RCS. This result should not be surprising, as humans are not rigid point targets but in fact have their own kinematics, namely, the periodic motion of the arms and legs, which is separate from gross motion along a path. Consider the classification criteria for the Swerling Models, summarized in Table 16-3. A signal is considered to be scan-to-scan decorrelated if the value measured over one coherent processing interval (CPI) is independent of values measured over previous CPIs. For typical radars, each CPI is composed of M pulses, so that the CPI may be computed as M times the PRI. If each individual pulse in a CPI results in an independent value for the RCS, then the signal is referred to as being pulse-to-pulse decorrelated. Because typical PRIs are on the order just fractions of a second, during the course of a single PRI slow-moving targets, such as humans, exhibit little fluctuation. Thus, one might expect that human targets would be scan-to-scan decorrelated.

The results in [79] verify this intuition. In particular, it was found that for low frequencies the RCS followed a Swerling III distribution, while for high frequencies the distribution more closely matched that of a Swerling I distribution. In both of these cases the decorrelation occurs scan-to-scan. Moreover, the transition between these two cases occurred between 1 and 2 GHz, although body position did affect the exact transition frequency. For example, the transition for a kneeling man occurred at a lower frequency than that of a standing man, who continued to exhibit Swerling III properties at 1.8 GHz. This too matches our intuition on human RCS, as when a human stands, the reflections from the torso act as a single dominant scatterer in addition to the many scatterers from other body parts. In the kneeling position, the torso is less prominent and thus on par with reflections from other body parts, which is consistent with the description matching a Swerling I target.

Although the modeling of human RCS is gaining increased attention [80-83], results so far are just beginning to provide an understanding of the RCS characteristics of humans. Much research remains to be performed to completely define and model human RCS, especially for the cases involving airborne applications, where elevation look angle has significant impact, as well as for a variety of target activities.

16.3 | SPECTROGRAM ANALYSIS OF HUMAN RETURNS

The spectrogram for a human target may be derived from the mathematical expression for the expected human radar return given in (16.3). First, the received data are organized into slow-time, fast-time matrix. *Fast-time* refers to the samples from each pulse within a

CPI. Each fast-time matrix entry is also known as a range bin, because of the relationship between sampling interval and range: $\Delta R = cT/2$, where T is time interval between transmitted pulses. *Slow-time* refers to samples taken from each CPI corresponding to the same range bin. Slow-time matrix entries are also known as Doppler bins, as taking a fast Fourier transform (FFT) across slow-time yields Doppler frequency. Once the slow-time, fast-time data matrix is formed, pulse compression across the fast-time dimension is performed so that a peak occurs at the range bin in which the target is present.

The pulse compression matched filter function is defined as the time-reversed replica of the transmitted signal, $h(t) = e^{-j\pi\gamma t^2}$. Thus, for a single body part

$$\begin{aligned} x(t) &= \int_{-\infty}^{\infty} s_r(u) h(t-u) du \\ &= \int_{t_d}^{t_d+\tau} a_t e^{j(-2\pi f_c t_d + \pi\gamma(u-t_d)^2)} e^{-j\pi\gamma(t-u)^2} du \end{aligned} \quad (16.21)$$

The peak occurs when $t = t_d$:

$$\begin{aligned} x_{peak} &= x(t_d) \\ &= \int_{t_d}^{t_d+\tau} a_t e^{j[-2\pi f_c t_d + \pi\gamma(u-t_d)^2]} e^{-j\pi\gamma(t_d-u)^2} du \\ &= \int_{t_d}^{t_d+\tau} a_t e^{-j2\pi f_c t_d} du \\ &= a_t \tau e^{-j2\pi f_c t_d} \end{aligned} \quad (16.22)$$

To find the pulse compressed output for the entire human return, it is only necessary to sum (16.22) for each body part and substitute the expression for time delay in terms of range

$$x_p[n] = \sum_{i=1}^{12} a_{t,i} \tau e^{-j\frac{4\pi f_c}{c} R_i[n]} \quad (16.23)$$

The spectrogram is then found by simply stacking the FFTs of short, overlapping time segments taken from the slow-time slice in (16.23). An example of the simulated spectrogram for a human target is shown in Figure 16-6. The strongest component of the return is caused by reflections from the torso, with its low-frequency, small-amplitude sinusoidal oscillatory motion. The larger amplitude oscillations are caused by reflections from the appendages, with the largest amplitude oscillation corresponding to the response from the feet—that is, the appendages that travel the farthest during the walking cycle.

It is important to note that the structure of the periodicities within the human spectrogram is unique to humans. Even the spectrograms of other animals are differentiable from human spectrograms. Consider the spectrograms for a human and a dog measured by Otero [85], which are shown in Figure 16-7. While the measured human spectrogram is similar to the simulated spectrogram in Figure 16-6, the spectrogram for a dog appears more noise-like as a result of the quadrupedal nature of the animal and equal-sized appendages. Indeed, Otero proposed a quantified approach to exploit these differences in

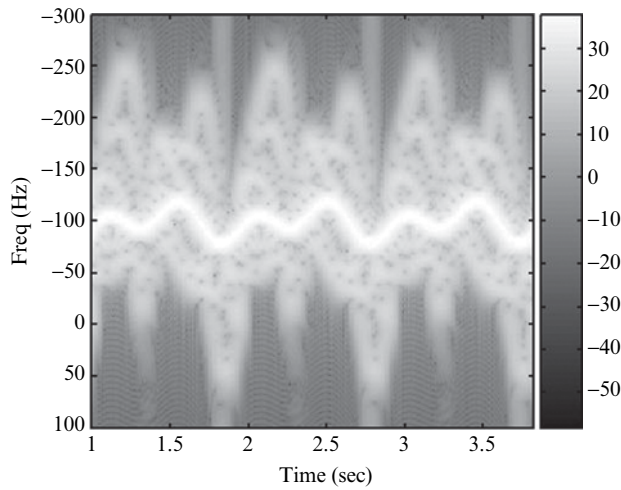


FIGURE 16-6 ■
Simulated
spectrogram for a
typical human target.
(From Gürbüz [84].
With permission.)

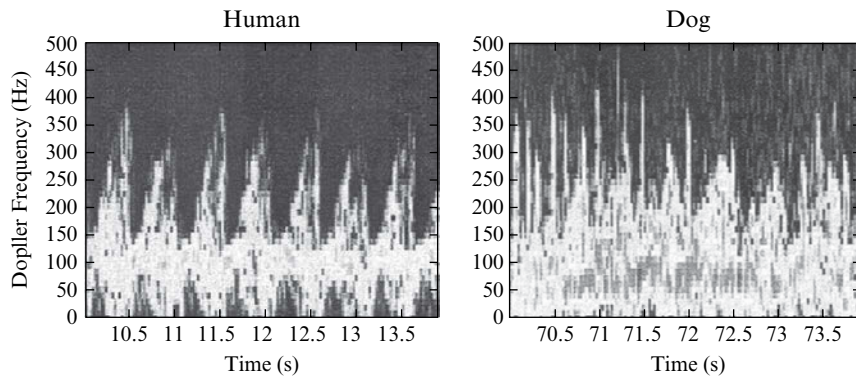


FIGURE 16-7 ■
Measured
spectrograms for a
human and dog.
(From Otero [85].
With permission.)

characteristics to analytically distinguish humans from dogs based on their spectrograms. In particular, the appendage-to-torso ratio (ATR) was proposed as a statistic that could be useful for such a purpose. The ATR is computed from what Otero terms a cadence frequency plot. This is basically the Fourier transform of the spectrogram, which exhibits peaks at frequencies corresponding to the constant-rate torso motion, the fundamental frequencies of appendage oscillations, and their harmonics. The amplitudes of these peaks are interpreted as being proportional to RCS. Thus, the ATR is defined as the ratio of the sum of the amplitudes of the peaks at the fundamental, second, and third harmonics to the amplitude of the torso peak:

$$ATR \equiv \frac{\sum_{n=1}^3 RCS_n}{RCS_0} \quad (16.24)$$

Otero postulated that the ATR for each animal could be unique and that, in particular, dogs would have a smaller ATR due to their significantly shorter legs relative to human legs.

Spectrogram analysis continues to be the primary technique for target characterization, discrimination, and classification today. Applications that have been investigated include gender discrimination [85, 86], suicide bomber detection [87], human activity and target classification [66, 88–91], gait analysis [92, 93], and through-the-wall detection [60].

16.4 | TECHNICAL CHALLENGES IN HUMAN DETECTION

The problem of human detection, however, is not just identifying whether a detected target is human but also determining whether a target is indeed present—a nontrivial task for human targets. As slow-moving targets, humans often fall below the MDV of GMTI radars. Moreover, their low RCS also makes human returns easily masked by ground clutter in many environments. Yet another challenge is that the RCS of the appendages is significantly lower than that of the torso. Finally, the nonlinear nature of the human return phase history also has a detrimental effect on the output SNR of typical radar detectors, leading to a decrease in detection performance when the target of interest is a human. Each of these effects will be examined in turn.

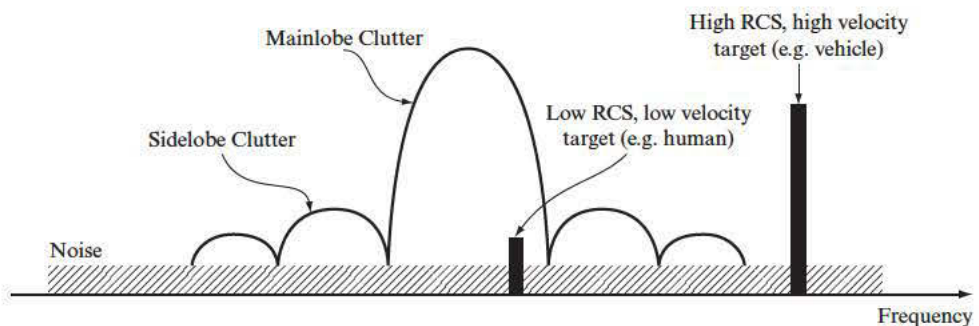
16.4.1 Slow-Moving Target Detection and GMTI Radars

Typical GMTI radars exploit Doppler shift to discriminate moving targets from stationary clutter, whose Doppler spread typically determines the MDV of GMTI radar. Slow-moving, low RCS targets, such as humans, however, possess very small Doppler shift and thus usually fall within the main lobe of the ground clutter (Figure 16-8), whose radar return is very strong and masks the human return. GMTI radars that attempt to detect targets with such small RCS and low Doppler often require expensive hardware and complicated signal processing [94]. The velocity synthetic aperture radar (VSAR) system [95], for example, reduces the MDV to 0.5 m/s but uses a large array and requires complex changes in hardware [96]. Recently, Tahmouh [97] proposed using micro-Doppler signatures to track dismounts moving in cross range with GMTI radar and showed that, despite their lower RCS, the side-to-side motion of humans and animals can make them easier to track in cross-range than vehicles. However, in high clutter environments, which are typical of many situations and applications such as those encountered with airborne radar, the use of STAP is a requirement to mitigate clutter and improve detection performance.

16.4.2 Mitigating the Effects of Clutter with STAP

STAP is an adaptive processing technique that suppresses interference while optimizing output SNR using operations in both the spatial and temporal domains. In a typical radar environment, there are three key sources of interference: (1) background noise from the radar receiver circuitry and other man-made sources and machinery; (2) clutter caused by

FIGURE 16-8 ■
Notional generic
Doppler spectrum
containing noise,
clutter, and target
components.



reflections from all objects in the scene except for the target, such as trees and buildings; and (3) electronic-countermeasures such as jammers used by hostile forces to impede radar performance. It can be shown that in angle-Doppler space the clutter falls along a straight line, typically referred to as the clutter ridge. STAP processing filters the data with a spatiotemporal filter so that the clutter ridge is nullified (i.e., canceled) or otherwise removed from the data. Note that since in the angle-Doppler domain the target does not lie on the clutter ridge, this operation does not adversely affect the target return; while in contrast the target signature increases in prominence.

The optimal STAP filter, discussed in detail in Chapter 10 on clutter suppression with STAP, can be determined as follows [73, 98, 99]. First, let us express the total interference, \mathbf{x}_n , as the sum of its main components

$$\mathbf{x}_n = \mathbf{w}_n + \mathbf{x}_c + \mathbf{x}_j \quad (16.25)$$

where \mathbf{w}_n is white noise, \mathbf{x}_c is clutter, and \mathbf{x}_j is jammer interference. As the optimum space-time filter is defined as the filter that maximizes the output signal-to-interference-plus-noise ratio (SINR), the total received signal, \mathbf{x} , is expressed as the sum of the target signal, \mathbf{s} , and interference plus noise: $\mathbf{x} = \mathbf{s} + \mathbf{x}_n$. After filtering with \mathbf{h} , the SINR may be written as

$$SINR = \frac{E[\mathbf{h}^H \mathbf{s} \mathbf{s}^H \mathbf{h}]}{E[\mathbf{h}^H \mathbf{x}_n \mathbf{x}_n^H \mathbf{h}]} = \frac{\mathbf{h}^H \mathbf{s} \mathbf{s}^H \mathbf{h}}{\mathbf{h}^H \mathbf{R}_I \mathbf{h}} \quad (16.26)$$

where, \mathbf{R}_I is the covariance matrix of interference, $\mathbf{R}_I \equiv E[\mathbf{x}_n \mathbf{x}_n^H]$, which is typically estimated by averaging over P range bins:

$$\hat{\mathbf{R}}_I = \frac{1}{P} \sum_{p=1}^P x_{n,p} x_{n,p}^{-1} \quad (16.27)$$

To maximize the SINR, first apply the Swartz inequality, which states that if $\boldsymbol{\alpha} = k\boldsymbol{\beta}$ for a constant scalar k , then

$$|\boldsymbol{\alpha}^H \boldsymbol{\beta}|^2 \leq \|\boldsymbol{\alpha}\|^2 \|\boldsymbol{\beta}\|^2 \quad (16.28)$$

Define $\boldsymbol{\alpha}^H \boldsymbol{\beta} = \mathbf{h}^H \mathbf{s}$, then $|\boldsymbol{\alpha}^H \boldsymbol{\beta}|^2 = \mathbf{h}^H \mathbf{s} \mathbf{s}^H \mathbf{h}$, which is the numerator of the expression for the SINR. To form one of the factors in the right-hand side of Schwartz's inequality such that it matches the denominator, first factor \mathbf{R}_I as the product of two matrices, namely, its square root: $\mathbf{R}_I = \mathbf{A}^H \mathbf{A}$. This in turn imposes the requirement that the interference covariance matrix be positive definite. Observe that the denominator can now be factored as

$$\mathbf{h}^H \mathbf{R}_I \mathbf{h} = \mathbf{h}^H \mathbf{A}^H \mathbf{A} \mathbf{h} = (\mathbf{A} \mathbf{h})^H \mathbf{A} \mathbf{h} = \|\mathbf{A} \mathbf{h}\|^2 \quad (16.29)$$

Thus, define $\boldsymbol{\alpha} = \mathbf{A} \mathbf{h}$, from which it follows that $\boldsymbol{\beta} = (\mathbf{A}^H)^{-1} \mathbf{s}$ to be consistent with our previous definition of the numerator.

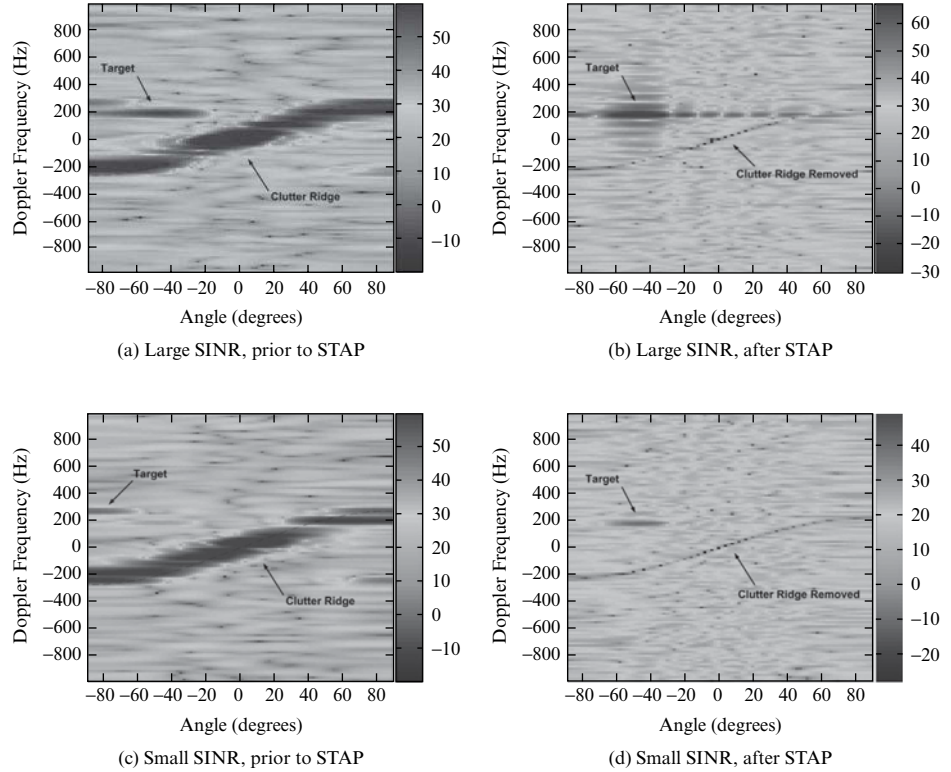
The Schwartz inequality then gives

$$\mathbf{h}^H \mathbf{s} \mathbf{s}^H \mathbf{h} \leq (\mathbf{h}^H \mathbf{R}_I \mathbf{h}) \left(\mathbf{s}^H \mathbf{R}_I^{-1} \mathbf{s} \right) \quad (16.30)$$

Rearranging to solve for SINR

$$SINR \leq \mathbf{s}^H \mathbf{R}_I^{-1} \mathbf{s} \quad (16.31)$$

FIGURE 16-9 ■
Illustration of
performance of
optimal STAP filter
on data for a human
target in clutter.



Equality occurs if and only if $\alpha = k\beta$ or $\mathbf{A}\mathbf{h}_{opt} = k(\mathbf{A}^H)^{-1}\mathbf{s}$. Solving for the optimum weight vector,

$$\mathbf{h}_{opt} = k\mathbf{A}^{-1}(\mathbf{A}^H)^{-1}\mathbf{s} = k(\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{s} = k\mathbf{R}_I^{-1}\mathbf{s} \quad (16.32)$$

The performance of (16.32) as a clutter cancellation filter is illustrated in Figure 16-9, which shows the angle-Doppler maps of a single human target before and after application of the optimal STAP filter. In Figures 16-9a and 16-9b, the SINR level of the human signal is large, so that it is clearly visible in the angle-Doppler maps. The clutter cancellation filter of (16.32) effectively removes the large clutter ridge in Figure 16-9a, leaving a clear target response in Figure 16-9b. However, as the interference levels increase and the SINR levels decrease, the target signature becomes less and less distinct in the data. Figures 16-9c and 16-9d show angle-Doppler maps for just such a case. Before clutter cancellation, the target signature is so weak due to masking by the clutter that it is barely discernable. After clutter cancellation, the clutter ridge is successfully removed, but the remaining target signal, although now visually observable, is still quite weak, limiting detection performance.

For airborne dismount detection applications, where the clutter levels are much higher and the larger distances between the target and the radar diminish the received signal power, it is much more likely to observe the high clutter scenario depicted in Figures 16-9c and 16-9d. Thus, while STAP processing is a necessary and critical component of any dismount detection system, STAP alone does not resolve all of the challenges to human detection, especially in the high clutter case.

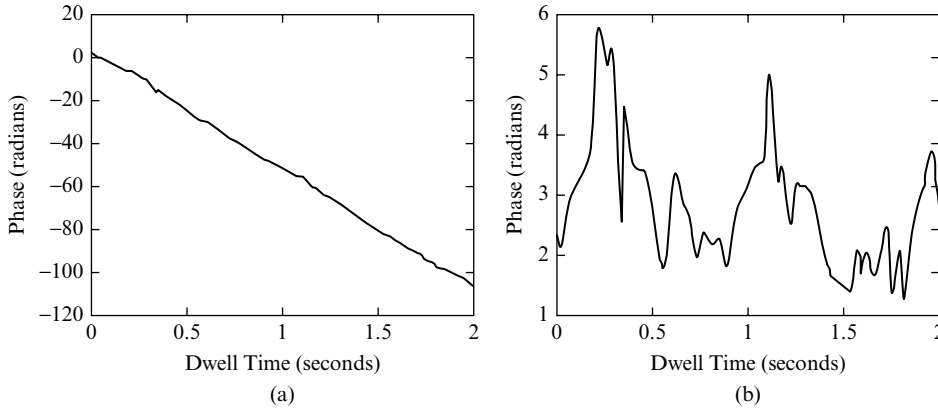


FIGURE 16-10 ■ Phase history of a typical human target walking radially away from the radar. (a) Total phase history. (From Gürbüz [84]. With permission.) (b) After linear component removed. (From Gürbüz et al. [100]. With permission.)

16.4.3 Inherent Output SNR Losses for Human Targets in Typical Detectors

The micro-Doppler effects that are evident in the human spectrogram can also be observed in the human phase history, which possesses a high degree of nonlinear phase, in contrast to that of point-target type targets, such as vehicles, whose phase history tends to be linear. Consider the phase history – the phase of the pulse compressed signal in (16.23) – of a typical human target walking radially away from the radar, as shown in Figure 16-10a. At first, the phase does not appear to be highly nonlinear; but if the linear component of the signal, which is caused by the total difference in radial velocity between the radar and the walking human, is removed, then as illustrated in Figure 16-10b, the nonlinear component of the human phase can be clearly observed. In fact, the remaining phase variation is so great that the nonlinear component of the signal spans the entire range between 0 and 2π .

Furthermore, the nonlinear nature of human phase history also results in significant losses in output SNR during the matched filtering stage of the receiver. Matched filtering is an operation used to boost the signal gain and is an integral component of the test statistic used in most detectors. However, typically linear-phase matched filters are used. Any type of Doppler filtering or Fourier transform processing is also a linear phase operation. For targets such as vehicles that can be modeled as point targets because they possess a linear phase history, these processors have a sufficiently good phase match that indeed the output SNR is improved. But when a significant phase mismatch is present—as is the case with linear-phase matched filtering of nonlinear phase human targets—the result is a drastic loss in output SNR inherent to the detection process itself.

The SNR loss incurred may be quantitatively analyzed as follows. Define the true target data as

$$\mathbf{s} = [\alpha_0 \quad \alpha_1 e^{j\theta_1} \quad \dots \quad \alpha_{N-1} e^{j(N-1)\theta_{N-1}}]^T \quad (16.33)$$

where N is the total number of pulses transmitted, and α_i and θ_i are generalized amplitude and phase factors, respectively. Note that in general both the amplitude and phase factors vary with slow time, n .

When \mathbf{s} is Doppler processed with a filter function \mathbf{w} , the output SNR is

$$SNR = \frac{E[\mathbf{w}^H \mathbf{s} \mathbf{s}^H \mathbf{w}]}{E[\mathbf{w}^H \mathbf{x}_n \mathbf{x}_n^H \mathbf{w}]} \quad (16.34)$$

where \mathbf{x}_n is the noise signal. For simplicity, assume that the noise has a covariance matrix of $\sigma_n^2 \mathbf{I}$. Then (16.34) may be simplified to

$$SNR = \frac{|\mathbf{w}^H \mathbf{s}|^2}{\sigma_n^2 |\mathbf{w}^H \mathbf{w}|} \quad (16.35)$$

The maximum output SNR is obtained when the signal \mathbf{s} is matched filtered against itself:

$$SNR_0|_{\mathbf{w}=\mathbf{s}} = \frac{1}{\sigma_N^2} \mathbf{s}^H \mathbf{s} \quad (16.36)$$

However, this would imply knowing the radar return for an unknown, not yet even detected target—a feat that is obviously impossible but nevertheless represents the greatest achievable performance and will thus be compared with the output SNR attained when the signal is matched filtered with a linear-phase filter:

$$\mathbf{w}_{LIN} = [\beta_1 \quad \beta_2 e^{j\phi} \cdots \beta_{N-1} e^{j(N-1)\phi}]^T \quad (16.37)$$

where β_i and ϕ are generalized amplitude and phase parameters, respectively. Here, the amplitude factor is left in a general form that varies with slow time, whereas only the phase has been restricted to be linear. The output SNR for such a linear phase filter is

$$SNR_{LIN}|_{\mathbf{w}=\mathbf{w}_{LIN}} = \frac{|\mathbf{w}_{LIN}^H \mathbf{s}|^2}{\sigma_N^2 \left| \sum_{n=0}^{N-1} \beta_n^2 \right|} \quad (16.38)$$

Thus, the SNR loss incurred from signal mismatch is

$$SNR_{Loss} = \frac{SNR_{LIN}}{SNR_0} = \frac{|\mathbf{w}_{LIN}^H \mathbf{s}|^2}{\left| \sum_{n=0}^{N-1} |\beta_n|^2 \right| \mathbf{s}^H \mathbf{s}} < 1 \quad (16.39)$$

There are three main factors that impact SNR loss: phase mismatch, amplitude mismatch, and dwell time. Figures 16-11a and 16-11b illustrate the effect of these factors by computing the SNR losses that would be incurred for the example phase history shown in Figure 16-10. An important observation is that the overall SNR loss is primarily due to phase mismatch, while amplitude mismatch has just a slight detrimental effect on output SNR.

Moreover, the SNR losses suffered can be quite large. In this example, the output SNR suffers from 20 to 30 dB of SNR loss, a significant loss that could mean failing to detect human targets in adverse clutter environments, where the target SNR is low. Furthermore, increasing the dwell time with the thought of exploiting signal integration to improve detection performance does not help in the case of human targets. Figure 16-11b shows the output SNR variation with dwell time for both the ideal, clairvoyant case, which assumes perfect matched filtering, and the linear-phase matched filter case. In the case of the ideal detector, the output SNR, and hence detection performance, does improve with dwell time. However, when phase mismatch is present, as is the case with linear-matched filtering, increasing the dwell time does not increase output SNR. This is a critical observation because it means that the inherent SNR loss problem cannot be overcome by simply collecting more data—instead, the detector itself must be redesigned to better

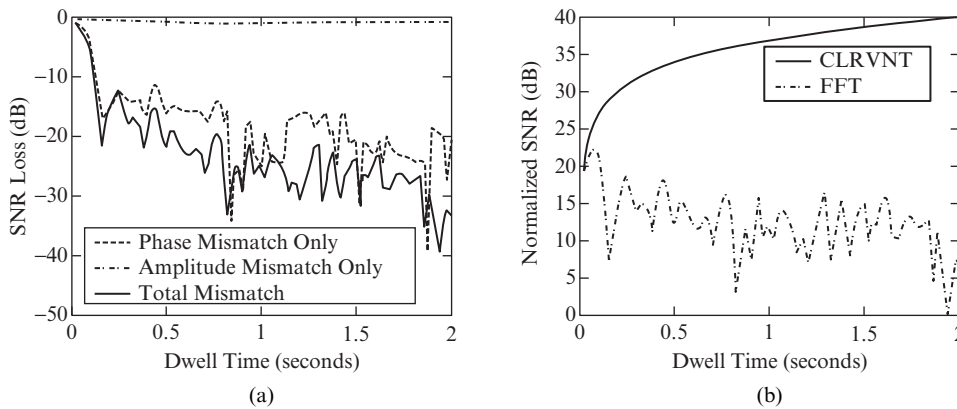


FIGURE 16-11 ■ (a) SNR loss variation over dwell time. (From Gürbüz [84]. With permission.) (b) Output SNR variation over dwell time normalized by input SNR for the ideal, clairvoyant detector and FFT-based linear-phase detector. (From Gürbüz et al. [100]. With permission.)

match the target of interest, in this case, human targets. This has led some researchers to propose incorporating a priori information about human kinematics into the detector design, as discussed in more detail in the next section.

16.5 | EXPLOITING KNOWLEDGE FOR DETECTION AND CLASSIFICATION

In practice, it is impossible to realize the ideal, clairvoyant detector because the exact target return is unknown. However, if the goal is to design detectors and classifiers specifically for human targets, then there is a priori knowledge about the target that can be exploited to design better algorithms—that is, it is known that the target is human. As mentioned during the discussion on human spectrograms, the kinematics of human gait are unique and represents a priori knowledge about the desired target that can be exploited for human detection and classification.

One example of incorporating kinematic information into detector design is to use the Boulic-Thalmann model to derive a parametric approximation to the expected target response [100]. Consider the antenna–target geometry illustrated in Figure 16-12, where it is assumed that human motion is linear along a constant angle, θ , relative to the initial target vector, r_1 . The vector h between the initial and final target locations represents

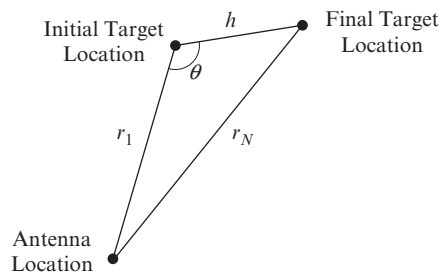


FIGURE 16-12 ■ Antenna–target geometry assuming human motion along a linear path. (From Gürbüz et al. [100]. With permission.)

human motion. Using the law of cosines

$$|\mathbf{r}_N|^2 = |\mathbf{r}_1|^2 + |\mathbf{h}|^2 - 2|\mathbf{r}_1||\mathbf{h}|\cos\theta \quad (16.40)$$

If it is further assumed that the overall distance traveled by the target is much less than the distance between the target and radar, $|\mathbf{h}| \ll |\mathbf{r}_1|$ and second-order terms are neglected, then

$$|\mathbf{r}_N| = |\mathbf{r}_1| \sqrt{1 + \left(\frac{|\mathbf{h}|}{|\mathbf{r}_1|}\right)^2 - 2\frac{|\mathbf{h}|}{|\mathbf{r}_1|}\cos\theta} \approx |\mathbf{r}_1| \left(1 - \frac{|\mathbf{h}|}{|\mathbf{r}_1|}\cos\theta\right) = |\mathbf{r}_1| - |\mathbf{h}|\cos\theta \quad (16.41)$$

Since the total SNR loss is caused primarily by phase mismatch, the received signal amplitude, $a_{t,i}$, is approximated as a constant, A/τ , with the range term in the amplitude approximated by the center of the range bin, r_b , in which the peak pulse compression output occurred. Thus, (16.23) may be simplified as

$$x_p[n] \approx \sum_{k=1}^{12} \frac{A_k}{r_b^2} e^{-j\frac{4\pi f_c}{c}(r_k - h_k \cos\theta_k)} \quad (16.42)$$

where $r = |\mathbf{r}_1|$, and $h = |\mathbf{h}|$. Observe that since the initial position of target as given by r is a constant, the only time-varying component of the phase is the factor containing the human motion term, h .

Suppose the human motion is modeled as a constant velocity point target. Then, $h = vTn$, where v is the target velocity, T is the PRI, and n is the number of transmitted pulses. In this case, the phase of x_p in (16.42) would just be the linear combination of linear phases and thus linear phase. As shown in the previous section, however, such a crude model leads to significant output SNR losses in the detector.

Another possible approximation would be to model the human motion by simply the torso motion, the dominant component in the human response. Using a simple sinusoidal approximation for the torso oscillations, approximate the human motion as a linear term, vTn , plus a sinusoidal factor. Then, x_p becomes

$$x_{p,\sin}[n] = \frac{A}{r_b^2} e^{j(Mn + C_1 + C_2 \cos(C_3 n + C_4))} \quad (16.43)$$

where M is the slope proportional to the Doppler frequency, C_1 is a factor that includes the range term r ; C_2 is the amplitude of the torso motion, C_3 represents the torso frequency, and C_4 is the torso phase. Since M and C_1 – C_4 are unknown model parameters, their values must be estimated from the received radar data via maximum likelihood estimation. These estimates can then be substituted back into (16.43) and used to match filter the data. Such sinusoidal-phase matched filters have been shown to achieve improvements in detection performance for human targets in preliminary studies [100, 101].

Further performance improvements may be achieved if the Boulic-Thalmann kinematic equations are used to compute the human motion term h_k for each body part. However, in this case, since the Boulic-Thalmann equations are not in closed form, each possible expected target response must be stored in a dictionary, or database, and matched filtered with the data. Sparse approximation techniques, such as orthogonal matching pursuit, can then be used to express the received radar data with the best possible linear combination of dictionary elements (possible target responses). Such an approach can

enable not only the design of a better matched filter, tailored to the expected human response, but also multiple target detection and a basis for classifying targets. For example, if the potential target response of other targets, such as vehicles or animals, is included in the dictionary, the optimal linear combination of dictionary also indirectly yields information on most likely target type [102].

The Boulic-Thalmann model has also been utilized in another knowledge-aided approach proposed by Bilik and Tabrikian [68]. In this method, physical models computed from the Boulic-Thalmann equations are stored in a database and used for classification between a single walking person, a pair of walking persons, and a slow-moving vehicle.

16.6 | SUMMARY

This chapter provided a brief overview of current systems and development efforts, kinematic and RCS models, technical challenges, and recently proposed novel ideas relating to human detection with radar. The expected radar return for a human target was analytically derived and the most widely used human kinematic model, the Boulic-Thalmann model, was described in detail. The concept of human micro-Doppler was introduced, along with the concept of human spectrograms and the application of spectrogram analysis for classification and identification of detected targets. Key technical challenges in human detection, such as the limitations of GMTI radars, STAP processors, and Fourier-based detectors, were presented, along with a brief introduction to knowledge-aided ideas that have been proposed as a solution to some of these issues.

There has been much progress in the field of human detection. However, there remains much work to be done to fully resolve questions in human modeling, kinematics, radar cross section, dismount detection, and target classification algorithm design. Advanced topics beyond the scope of this chapter but that also constitute important applications of human radar detection include the following:

- Through-the wall detection of humans, which aims at using ultra wideband radar to detect human respiration, for important applications, such as indoor surveillance and human search and rescue of fire and earthquake victims [60, 103–106]
- Synthetic aperture radar detection and imaging of human targets [94, 107, 108]
- Radar networks [109–112] for border security and surveillance.

16.7 | FURTHER READING

Currently there is no text book solely dedicated to the subject of human detection with radar. However, there are a number of good texts on key topics relevant to human detection. The text by Schleher [113] provides a solid introduction to moving target indication and pulse-Doppler radar. A solid introduction to STAP and related advanced topics can be found in the books by Klemm [99, 114]. For details on the detection and estimation theory that forms the foundation for human detection with radar, there are several good textbooks available, including those by Van Trees [115–118], Kay [119, 120], and Haykin [121]. The text by Chen [122] provides detailed explanation of the physics and mathematics of the micro-Doppler effect, giving examples using inanimate objects—such as an oscillating pendulum, rotating helicopter or wind-turbine blades—as well as living beings, such as a

person walking with swinging arms and legs, a flying bird, and quadruped animals. Much more information directly related to human detection and classification algorithms, and their applications, may be found in individual papers published in journals and conference proceedings. Good conferences to look at include the *IEEE Radar Conference* and *SPIE Remote Sensing, and Security and Defense Conferences*, published in the *SPIE Proceedings*. Journals in which key papers have been published include the *IEEE Transactions on Aerospace and Electronics Systems*, and the *IET Journal on Radar, Sonar and Navigation*. Interesting summary papers on related topics, such as gait analysis and STAP, have also appeared in the *IEEE Signal Processing Magazine*.

16.8 | REFERENCES

- [1] Benezeth, Y., Emile, B., Laurent, H., and Rosenberger, C., "Vision-Based System for Human Detection and Tracking in Indoor Environment," *International Journal of Social Robotics*, vol. 2, no. 1, 2010, pp. 41–52.
- [2] Geronimo, D., Lopez, A.M., Sappa, A.D., and Graf, T., "Survey of Pedestrian Detection for Advanced Driver Assistance Systems," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1239–1258, July 2010.
- [3] Boulton, T.E., Michaels, R.J., Gao, X., and Eckmann, M., "Into the Woods: Visual Surveillance of Noncooperative and Camouflaged Targets in Complex Outdoor Settings," in *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1382–1402, October 2001.
- [4] Niu, W., Long, J., Han, D., and Wang, Y., "Human Activity Detection and Recognition for Video Surveillance," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, vol. 1, pp. 719–722, June 2004.
- [5] Yoon, S.M. and Kim, H., "Real-Time Multiple People Detection Using Skin Color, Motion and Appearance Information," in *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*, pp. 331–224, September 2004.
- [6] Choi, Y.K., Kim, K.M., Jung, J.W., Chun, S.Y., and Park, K.S., "Acoustic Intruder Detection System for Home Security," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 130–138, February 2005.
- [7] Shoji, Y., Takasuka, T., and Yasukawa, H., "Personal Identification Using Footstep Detection," in *Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 43–47, November 2004.
- [8] Ekimov, A.E. and Sabatier, J.M., "A Review of Human Signatures in Urban Environments Using Seismic and Acoustic Methods," In *Proceedings of the IEEE Conference on Technologies for Homeland Security*, Waltham, MA, pp. 215–220, May 12–13, 2008.
- [9] Mazarakis, G.P. and Avaritsiotis, J.N., "A Prototype Sensor Node for Footstep Detection," in *Proceedings of the Second European Workshop on Wireless Sensor Networks*, pp. 415–418, 2005.
- [10] Succi, G., Clapp, D., Gampert, R., and Prado, G., "Footstep Detection and Tracking," in *Proceedings of the SPIE*, vol. 4393, pp. 22–29, 2001.
- [11] Richman, M.S., Deadrick, D.S., Nation, R.J., and Whitney, S.L., "Personnel Tracking Using Seismic Sensors," in *Proceedings of the SPIE*, vol. 4393, 2001.
- [12] Benezeth, Y., Emile, B., Laurent, H., and Rosenberger, C., "A Real Time Human Detection System Based on Far Infrared Vision," in *Proceedings of the 3rd International Conference on Image and Signal Processing*, Cherbourg-Octeville, France, pp. 76–84, July 1–3, 2008.

- [13] Broggi, A., Fascioli, A., Carletti, M., Graf, T., and Meinecke, M., "A Multi-Resolution Approach for Infrared Vision-Based Pedestrian Detection," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 7–12, June 2004.
- [14] Linzmeier, D.T., Skutnek, M., Mekhaie, M., and Dietmayer, K.C.J., "A Pedestrian Detection System Based on Thermopile and Radar Sensor Data Fusion," in *Proceedings of the 7th International Conference on Information Fusion*, vol. 2, pp. 1272–1279, July 2005.
- [15] Cui, J., Zha, H., Zhao, H., and Shibasaki, R., "Laser-Based Detection and Tracking of Multiple People in Crowds," *Computer Vision and Image Understanding*, vol. 106, no. 2–3, pp. 300–312, May–June 2007.
- [16] Milch, S. and Behrens, M., "Pedestrian Detection with Radar and Computer Vision," in *Proceedings of the Conference on Progress in Automobile Lighting*, Darmstadt, Germany, 2001.
- [17] Jane's Radar and Electronic Warfare Systems Online. Available at <http://www.janes.com>.
- [18] Global Security, "AN/PPS-5B Ground Surveillance Radar Set." Available at <http://www.globalsecurity.org>.
- [19] Military Periscope Database, "AN/PPS-5 MSTAR Ground Radar." Available at <http://periscope.ucg.com>.
- [20] SRCtech, "SR Hawk Info Sheet." Available at <http://www.srctecinc.com>.
- [21] ICX Technologies, "STS-700, STS-1400, STS-2800 Mid-Range Surveillance Radars." Available at <http://www.icxt.com>.
- [22] ICX Technologies, "STS-4400 Rugged Wide Area Surveillance Radar." Available at <http://www.icxt.com>.
- [23] ICX Technologies, "STS-12000 Remote Detection and Tracking Sensor with Fast Scan Ability." Available at <http://www.icxt.com>.
- [24] Tsunoda, S.I., Pace, F., Stence, J., Woodring, M., Hensley, W.H., Doerry, A.W., and Walker, B.C., "Lynx: A High-Resolution Synthetic Aperture Radar," *SPIE Aerospace*, vol. 3704, 1999.
- [25] "Testing Complete on UAS Equipped with Dismount Detection Radar," Defense File, May 25, 2010.
- [26] "ER/MP Gray Eagle: Enhanced MQ-1C Predators for the Army," *Defense Industry Daily*. Available at <http://www.defenseindustrydaily.com/awrior-emp-on-enhanced-predator-for-the-army-03056/>.
- [27] Northrop Grumman AN/ZPY-1 STARlite Small Tactical Radar – Lightweight. Available at <http://www.es.northropgrumman.com/solutions/starlite/assets/starlite.pdf>.
- [28] Hodge, N., "U.S. Military Turns to 'Vader' to Hunt Rebel Scum in Afghanistan," *Wired Magazine*, October 28, 2009.
- [29] Jane's Electronic Mission Aircraft Online. Available at <http://www.janes.com>.
- [30] "Man-Tracking Radar to Fight Afghan IEDs," *Aviation Week*, February 17, 2010.
- [31] "Man-Hunting Radar," *Defense News*, April 26, 2010.
- [32] "AACER System Would Detect Ground Targets while Passing Data about Them," *SpaceWar Express*, November 2, 2005.
- [33] Robinson, C.A., "Radar Counters Camouflage," *Signal Online*, June 2007.
- [34] "A160T flies FORESTER in Belize," *Aviation Week Defense Blog*, September 8, 2010. Available at <http://www.aviationweek.com>.
- [35] Novacheck, T.F., "The Biomechanics of Running," *Gait and Posture*, vol. 7, 1998, pp. 77–95.

- [36] Behncke, H., "A Mathematical Model for Competitive Running," *Methods of Operation Research*, no. 49, 1985, pp. 471–498.
- [37] Watanabe, K., "Kinematical Analysis and Measurement of Sports Form," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 36, no. 3, May 2006.
- [38] Orendurff, M.S., Segal, A.D., Berge, J.S., Flick, K.C., Spanier, D., and Klute, G.K., "The Kinematics and Kinetics of Turning: Limb Asymmetries Associated with Walking a Circular Path," *Gait and Posture*, vol. 23, pp. 106–111, 2006.
- [39] Smith, B., Ashton, K.M., Bohl, D., Clark, R.C., Metheny, J.B., and Klassen, S., "Influence of Carrying a Backpack on Pelvic Tilt, Rotation and Obliquity in Female College Students," *Gait and Posture*, vol. 23, pp. 263–267, 2006.
- [40] Fowler, N.E., Rodacki, A.L.F., and Rodacki, C.D., "Changes in Stature and Spine Kinematics during a Loaded Walking Task," *Gait and Posture*, vol. 23, pp. 133–141, 2006.
- [41] Fiolokowski, P., Horodyski, M., Bishop, M., Williams, M., and Stylianou, L., "Changes in Gait Kinematics and Posture with the Use of a Front Pack," *Ergonomics*, vol. 49, no. 9, pp. 885–894, July 15, 2006.
- [42] Boulgouris, N.V., Hatzinakos, D., and Plataniotis, K.N., "Gait Recognition: A Challenging Signal Processing Technology for Biometric Identification," *IEEE Signal Processing Magazine*, November 2005.
- [43] Little, J.J. and Boyd, J.E., "Recognizing People by Their Gait: The Shape of Motion," *Videre: Journal of Computer Vision Research*, vol. 1, no. 2, Winter 1998.
- [44] Liu, L.F., Jia, W., and Zhu, Y.H., "Survey of Gait Recognition," *Lecture Notes in Computer Science: Emerging Intelligent Computing Technology and Applications with Aspects of Artificial Intelligence*, vol. 5755, pp. 652–659, 2009.
- [45] Chen, V.C., "Analysis of Radar Micro-Doppler Signature with time-Frequency Transform," in *Proceedings of the 10th IEEE Workshop on Statistical and Array Processing*, pp. 463–466, 2000.
- [46] Chen, V.C., Li, F., Ho, S.S., and Wechsler, H., "Micro-Doppler Effect in Radar: Phenomenon, Model and Simulation Study," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 2–21, January 2006.
- [47] Zhang, Z., Pouliquen, P., Waxman, A., and Andreou, A., "Acoustic Micro-Doppler Radar for Human Gait Imaging," *Journal of the Acoustic Society of America Express Letters*, vol. 121, no. 3, pp. 110–113, 2007.
- [48] Weir, R.F. and Childress, D.S., "A New Method of Characterizing Gait Using a Portable, Real-Time, Ultrasound Ranging Device," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 4, pp. 1810–1812, October 30–November 2, 1997.
- [49] Weir, R.F. and Childress, D.S., "A Portable, Real-Time, Clinical Gait Velocity and Analysis System," *IEEE Transactions on Rehabilitation Engineering*, vol. 5, no. 4, pp. 310–321, 1997.
- [50] Geisheimer, J.L., Marshall, W.W., and Greneker, E., "A Continuous-Wave (CW) Radar for Gait Analysis," *Signals, Systems and Computers*, vol. 1, pp. 834–838, 2001.
- [51] van Dorp, P. and Groen, F.C.A., "Human Walking Estimation with Radar," *IEE Proceedings on Radar, Sonar, and Navigation*, vol. 150, no. 5, pp. 356–365, October 2003.
- [52] G.E. Smith, K. Woodbridge, and C.J. Baker, "Multistatic micro-Doppler signature of personnel," in *Proc. IEEE Radar Conference*, 26–30 May, 2008, pp. 1–6.

- [53] G.E. Smith, K. Woodbridge, C.J. Baker, H. Griffiths, "Multistatic micro-Doppler radar signatures of personnel targets," *IET Journal of Signal Processing*, Vol. 4, No. 3, June 2010, pp. 224–233.
- [54] S.S. Ram, and H. Ling, "Microdoppler signature simulation of computer animated human and animal motions," *IEEE International Symposium on Antennas and Propagation*, 2008, pp. 1–4.
- [55] S.S. Ram, and H. Ling, "Simulation of human microDopplers using computer animation data," *IEEE Radar Conference*, 2008, pp. 1–6.
- [56] Vignaud, L., Ghaleb, A., and Le Kernec, J., "Radar High Resolution Range and Micro-Doppler Analysis of Human Motions," in *Proceedings of the IEEE International Radar Conference*, Bordeaux, France, October 12–16, 2009.
- [57] Yang, Y., Zhang, W., and Lu, C., "Classify Human Motions Using Micro-Doppler Radar," in *Proceedings of SPIE*, vol. 6944, 2008.
- [58] The Motion Research Laboratory, Carnegie Mellon University. Available at <http://mocap.cs.cmu.edu/>.
- [59] S.S. Ram, R. Bhalla, and H. Ling, "Simulation of human radar signatures in the presence of ground," *IEEE International Symposium on Antennas and Propagation*, 2009, pp. 1–4.
- [60] S.S. Ram, C. Christianson, K. Youngwook, and H. Ling, "Simulation and analysis of human micro-dopplers in through-wall environments," *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 48, No. 4, Part 2, 2010, pp. 2015–2023.
- [61] Geisheimer, J.L., Greneker, E.F., and Marshall, W.S., "A High-Resolution Doppler Model of Human Gait," in *Proceedings of SPIE, Radar Sensor Technology and Data Visualization*, vol. 4744, 2002.
- [62] Raj, R.G., Chen, V.C., and Lipps, R., "Analysis of Radar Dismount Signatures via Non-parametric and Parametric Models," in *Proceedings of the IEEE Radar Conference*, 2009.
- [63] Raj, R.G., Chen, V.C., and Lipps, R., "Analysis of Radar Human Gait Signatures," *IET Signal Processing*, vol. 4, no. 3, pp. 234–244, 2010.
- [64] He, F., Huang, X., Liu, C., Zhou, Z., and Fan, C., "Modeling and Simulation Study on Radar Doppler Signatures of Pedestrian," in *Proceedings of the IEEE Radar Conference*, Washington, DC, pp. 1322–1326, May 10–14, 2010.
- [65] Zhang, Y., "Research on Radar Feature of Human Micro-Motion," *National University of Defense Technology*, Changsha, China, 2009. (in Chinese)
- [66] Boulic, R., Thalmann, M.N., and Thalmann, D., "A Global Human Walking Model with Real-Time Kinematic Personification," *Visual Computing*, vol. 6, pp. 344–358, 1990.
- [67] van Dorp, P. and Groen, F.C.A., "Feature-Based Human Motion Parameter Estimation with Radar," *IET Proc. on Radar, Sonar, and Navigation*, vol. 2, no. 2, pp. 135–145, 2008.
- [68] Bilik, I. and Tabrikian, J., "Radar Target Classification Using Doppler Signatures of Human Locomotion Models," *IEEE Transactions on AES*, vol. 43, no. 4, pp. 1510–1522, October 2007.
- [69] Gürbüz, S.Z., Melvin, W.L., and Williams, D.B., "Detection and Identification of Human Targets in Radar Data," in *Proceedings of SPIE Defense and Security Symposium*, Orlando, FL, April 9–13, 2007.
- [70] Hersey, R.K., Melvin, W.L., and Culpepper, E., "Dismount Modeling and Detection from Small Aperture Moving Radar Platforms," in *Proceedings of the IEEE Radar Conference*, 2008.

- [71] The RoyMech Website. Available at http://www.roymech.co.uk/Useful_Tables/Human/Human_sizes.html.
- [72] Yamada, N., Tanaka, Y., and Nishikawa, K., "Radar Cross Section for a Pedestrian in 76 GHz Band," in *Proceedings of the European Microwave Conference*, vol. 2, October 4–6, 2005.
- [73] Richards, M.A., *Fundamentals of Radar Signal Processing*, McGraw Hill, New York, 2005.
- [74] R. Linnehan, R. Deming, and J. Schindler, "Multipath analysis of dismount radar responses," in *Proc. IEEE Radar Conference*, 2011, pp. 474–479.
- [75] P. Setlur, M. Amin, and F. Ahmad, "Multipath Doppler signatures from targets moving behind walls," in *Proc. IEEE Radar Conference*, 2010, pp. 799–803.
- [76] Knott, E.F., Shaeffer, J.F., and Tuley, M.T., *Radar Cross Section*, 2d ed., Artech House, Boston, MA, 1993.
- [77] Ruck, G.T., Barrick, D.E., Stuart, W.D., and Krichbaum, C.K., *Radar Cross Section Handbook*, vol. 1, Plenum Press, London, 1970.
- [78] Trott, K.D., "Stationary Phase Derivation for RCS of an Ellipsoid," *IEEE Antennas and Wireless Propagation Letters*, vol. 6, pp. 240–243, 2007.
- [79] Dogaru, T., Nguyen, L., and Le, C., "Computer Models of the Human Body Signature for Sensing Through the Wall Radar Applications," U.S. Army Research Laboratory Technical Report ARL-TR-4290, September 2007.
- [80] Dogaru, T. and Nguyen, L., "FDTD Models of Electromagnetic Scattering by the Human Body," in *Proceedings of the IEEE Antennas and Propagation Society International Symposium*, 2006.
- [81] Stratton, S.R. and Bender, R.L., "Radar Cross-Section (RCS) Measurements of a Dismount with Rocket-Propelled Grenade (RPG) Launcher at Ka-Band," U.S. Army Research Laboratory Technical Report ARL-TR-3855, July 2006.
- [82] Le, C. and Dogaru, T., "Numerical Modeling of the Airborne Radar Signature of Dismount Personnel in the UHF-, L-, and Ka-Bands," U.S. Army Research Laboratory Technical Report ARL-TR-4336, December 2007.
- [83] Tahmoush, D. and Silvius, J., "Simplified Model of Dismount Micro-Doppler and RCS," in *Proceedings of the IEEE Radar Conference*, 2010.
- [84] Gürbüz, S.Z., "Radar Detection and Identification of Human Signatures Using Moving Platforms," PhD diss., Georgia Institute of Technology, Department of Electrical and Computer Engineering, December 2009.
- [85] Otero, M., "Application of a Continuous Wave Radar for Human Gait Recognition," in *Proceedings of SPIE, Signal Processing, Sensor Fusion and Target Recognition XIV*, vol. 5809, pp. 538–548, 2005.
- [86] Tahmoush, D. and Silvius, J., "Radar microDoppler for Security Applications: Modeling Men versus Women," in *Proceedings of the Antennas and Propagation Society International Symposium*, Charleston, SC, June 1–5, 2009.
- [87] G. Greneker, "Very low cost stand-off suicide bomber detection system using human gait analysis to screen potential bomb carrying individuals," in *Proc. SPIE*, Vol. 5788, 2005.
- [88] Kim, Y. and Ling, H., "Human Activity Classification Based on Micro-Doppler Signatures Using an Artificial Neural Network," in *Proceedings of the Antennas and Propagation Society International Symposium*, 2008.
- [89] Yang, Y., Lei, J., Zhang, W., and Lu, C., "Target Classification and Pattern Recognition Using Micro-Doppler Radar Signatures," in *Proceedings of the 7th ACIS International*

Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, pp. 213–217, 2006.

- [90] Tan, R. and Bender, R., “Analysis of Doppler Measurements of People,” in *Proceedings of SPIE, Targets and Backgrounds XII: Characterization and Representation*, vol. 6239, 623908-1, 2006.
- [91] Lyonnet, B., Ioana, C., and Amin, M.G., “Human Gait Classification Using Micro-Doppler Time-Frequency Signal Representations,” in *Proceedings of the IEEE Radar Conference*, 2010.
- [92] Tivive, F.H.C., Bouzerdoun, A., and Amin, M.G., “A Human Gait Classification Method Based on Radar Doppler Spectrograms,” in *EURASIP Journal on Advances in Signal Processing*, Article 389716, 2010.
- [93] Hornsteiner, C. and Detlefsen, J., “Characterisation of Human Gait Using a Continuous-Wave Radar at 24 GHz,” *Advances in Radio Science*, vol. 6, pp. 67–70, 2008.
- [94] Goldstein, J.S., Picciolo, M.L., Rangaswamy, M., and Griesbach, J.D., “Detection of Dismounts Using Synthetic Aperture Radar,” in *Proceedings of the IEEE Radar Conference*, 2010.
- [95] Friedlander, B. and Porat, B., “VSAR: A High Resolution Radar System for Detection of Moving Targets,” *IEEE Proceedings of the Radar, Sonar, and Navigation*, vol. 144, no. 4, pp. 205–218, 1997.
- [96] Linnehan, R., Perlovsky, L., Mutz, C., Rangaswamy, M., and Schindler, J., “Detecting Multiple Slow-Moving Targets in SAR Images,” in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2004.
- [97] Tahmoush, D., Clark, J., and Silvius, J., “Tracking of Dismounts Moving in Cross-Range Using GMTI Radar,” in *Proceedings of the International Radar Symposium*, 2010.
- [98] Ward, J., “Space-Time Adaptive Processing for Airborne Radar,” *Lincoln Laboratory Technical Report ESC-TR-94-109*, December 1994.
- [99] Klemm, R., *Principles of Space-Time Adaptive Processing*, 3d ed., IET Publishing, 2006.
- [100] Gürbüz, S.Z., Melvin, W.L., and Williams, D.B., “A Non-linear Phase, Model-Based Human Detector for Radar,” *IEEE Trans. Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 2502–2513, October 2011.
- [101] S.Z. Gürbüz, W.L. Melvin, and D.B. Williams, “Comparison of radar-based human detection techniques,” 41st Asilomar Conference, Monterey, CA, November 4–7, 2007.
- [102] Gürbüz, S.Z., Melvin, W.L., and Williams, D.B., “Radar-Based Human Detection via Orthogonal Matching Pursuit,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2778–2781, 2010.
- [103] Greneker, E.F., “Radar Flashlight for through the Wall Detection of Humans,” in *Proceedings of SPIE, Targets and Backgrounds: Characterization and Representation IV*, vol. 3375, pp. 280–285, April 1998.
- [104] Ram, S.S., Li, Y., Lin, A., and Ling, H., “Doppler-Based Detection and Tracking of Humans in Indoor Environments,” *Journal of the Franklin Institute*, vol. 345, pp. 679–699, 2008.
- [105] Zaikov, E., Sachs, J., Aftanas, M., and Rovnakova, J., “Detection of Trapped People by UWB Radar,” in *Proceedings of the German Microwave Conference*, 2008.
- [106] Bugaev, A.S., Vasilev, A., Ivashov, S.I., and Chapurskii, V.V., “Radar Methods of Detection of Human Breathing and Heartbeat,” *Journal of Communications Technology and Electronics*, vol. 51, no. 10, pp. 1154–1168, 2006.

- [107] Chapin, E. and Chen, C.W., “Preliminary Results from an Airborne Experiment Using Along-Track Interferometry for Ground Moving Target Indication,” in *Proceedings of the IEEE International Radar Conference*, May 9–12, 2005.
- [108] Shingu, G., Takizawa, K., and Ikegami, T., “Human Body Detection Using MIMO-UWB Radar Sensor Network in an Indoor Environment,” in *Proceedings of the 9th International Conference on Parallel and Distributed Computing, Applications and Technologies*, pp. 437–442, 2008.
- [109] Nohara, T.J., Weber, P., Jones, G., Ukrainec, A., and Premji, A., “Affordable High-Performance Radar Networks for Homeland Security Applications,” in *Proceedings of the IEEE Radar Conference*, 2008.
- [110] Arora, A., Dutta, P., Bapat, S., Kulathumani, V., Zhang, H., Naik, V., et. al., “A Line in the Sand: A Wireless Sensor Network for Target Detection, Classification, and Tracking,” *Computer Networks*, vol. 46, no. 5, pp. 605–634, December 5, 2004.
- [111] Guo, H., Coetzee, S., Mason, D., Woodbridge, K., and Baker, C., “Passive Radar Detection Using Wireless Networks,” in *Proceedings of the IET International Conference on Radar Systems*, Edinburgh, UK, October 15–18, 2007.
- [112] Lane, R.O. and Hayward, S.D., “Detection Personnel in Wooded Areas Using MIMO Radar,” in *Proceedings of the IET International Conference on Radar Systems*, Edinburgh, UK, October 15–18, 2007.
- [113] Schleher, D.C., *MTI and Pulsed Doppler Radar with MATLAB*, 2d ed., Artech House, Boston, MA, 2009.
- [114] Klemm, R., *Applications of Space-Time Adaptive Processing*, IET Publishers, 2004.
- [115] Van Trees, H.L., *Detection, Estimation, and Modulation Theory, Part I*, Wiley-Interscience, 2001.
- [116] Van Trees, H.L., *Detection, Estimation, and Modulation Theory, Part II: Nonlinear Modulation Theory*, Wiley-Interscience, 2002.
- [117] Van Trees, H.L., *Detection, Estimation, and Modulation Theory, Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise*, Wiley-Interscience, 2001.
- [118] Van Trees, H.L., *Detection, Estimation, and Modulation Theory, Part IV: Optimum Array Processing*, Wiley-Interscience, 2002.
- [119] Kay, S.M., *Fundamentals of Statistical Signal Processing, vol. I: Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [120] Kay, S.M., *Fundamentals of Statistical Signal Processing, vol. II: Detection Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1998.
- [121] Haykin, S.S. and Steinhardt, A., *Adaptive Radar Detection and Estimation*, Wiley-Interscience, 1992.
- [122] Chen, V.C., *The Micro-Doppler Effect in Radar*, Artech House, Boston, MA, 2011.

16.9 | PROBLEMS

1. For simplicity, consider only the torso response of a human target, as defined by equations (16.11)–(16.13). Compute and plot the trajectories. How does changing RV affect the torso trajectory?
2. Compute the torso component of the pulse compressed human return, as defined in (16.23), for an antenna located at an elevation of 1000 m and distance of 1000 m from

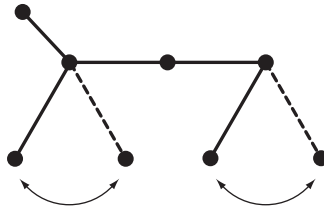


FIGURE 16-13 ■
Pendulum-based
dog model.

a man walking away from the antenna at a speed of 2 m/s. Assume the man's thigh height is 1 m and that the radar operates at a center frequency of 10 GHz, and pulse repetition interval of 0.2 ms. Remove the linear component of the phase history to isolate the nonlinear component. How does changing the dwell time affect the degree of nonlinearity of the human phase history?

3. For the target profile described in problem 2, compute the SNR loss (16.39) due to phase mismatch for varying dwell times. How does changing the dwell time affect the amount of SNR loss caused by phase mismatch with the detector?
4. Compute a sinusoidal approximation to the torso response, as given in (16.42), for the target described in problem 2. How does the SNR loss of the sinusoidal detector compare with that of the linear detector?
5. Compute the spectrogram of the torso response for the human target specified in problem 2.
6. Consider a pendulum-based model for a medium-sized dog, in which the body is modeled as being composed of eight point targets (Figure 16-13) and legs move as a synchronized pendulum oscillating with frequency ω . Assume the neck is 10 cm long, the body is 30 cm long, and each leg is 20 cm long. Compute the spectrogram based on this dog model for a radar with a center frequency of 10 GHz, and pulse repetition interval of 0.2 ms.
7. How does the modeled dog spectrogram obtained in problem 6 compare with the measured dog spectrogram given in Figure 16-7? Is the pendulum-based dog model a good model? How could the model be improved?
8. Compare the modeled dog spectrogram obtained in problem 6 with that of the human torso response and overall human signature shown in Figure 16-7. What differences do you observe?
9. Compute the best fitting sinusoidal detector for the dog model data generated in problem 6. How does the performance of the sinusoidal detector compare with that of a linear detector?
10. Compute the SNR loss for a single response containing three people of varying sizes (see Table 16-2). Does the SNR loss increase as more human targets fall within a single range bin?

Advanced Processing Methods for Passive Bistatic Radar Systems

P. Lombardo, F. Colone

Chapter Outline

17.1	Introduction	739
17.2	Evaluation of the 2D-CCF for the Passive Radar Coherent Integration	747
17.3	Direct Signal and Multipath/Clutter Cancellation Techniques	755
17.4	Signal Processing Techniques for Reference Signal Cleaning and Reconstruction	766
17.5	2D-CCF Sidelobe Control	775
17.6	Multichannel Processing for Detection Performance Improvement	791
17.7	Summary	814
17.8	Acknowledgments	815
17.9	Further Reading	815
17.10	References	815
17.11	Problems	819

17.1 INTRODUCTION

Recently, passive bistatic radar (PBR) has received renewed interest for surveillance purposes. It allows for target detection and localization with many advantages such as low cost, covert operation, low vulnerability to electronic countermeasure, and reduced electromagnetic pollution. PBR exploits an existing transmitter as an illuminator of opportunity. A receiving system is appropriately designed to receive the echoes reflected from air targets and sited to provide coverage of a specific area [1].

Numerous transmitters for telecommunications, radio navigation, and remote sensing applications can be used as sources of opportunity [2, 3], including:

- Broadcast transmitters of commercial radio stations in the frequency modulation (FM) radio band 88–108 MHz are especially attractive for the generally high level of transmitted power, for the wide coverage, and for the limited cost of the required receivers [4, 5].
- Digital video broadcasting-terrestrial (DVB-T) and digital audio broadcasting (DAB) transmitters are replacing their analog counterparts in a large portion of the world and are characterized by wide coverage and wider bandwidth (and thus better resolution) than FM signals. In addition, they benefit from the typical characteristics of their digital orthogonal frequency-division multiplexing (OFDM) modulation, together with their known coding scheme [6–10].

- Transmitters in the high frequency (HF) band for the potential capability to detect targets at very long ranges, including amplitude modulation (AM) radio, digital radio mondiale (DRM), and existing frequency modulated continuous-wave (FMCW) over-the-horizon (OTH) radar illuminators [11, 12].
- Transmitters for base stations of Global System for Mobile Communications (GSM), Universal Mobile Telecommunications System (UMTS), Worldwide Interoperability for Microwave Access (WiMAX), and forthcoming generation, such as Long Term Evolution (LTE), of mobile personal communication and network connection provide a largely populated and well-connected network of sources of opportunity, with medium bandwidths [13-15].
- WiFi transmitters provide a very local but potentially wide bandwidth and well-controlled signals useful when aiming at indoor surveillance or at monitoring small external areas [16-21].
- Global navigation satellite systems, like GPS, GLObal NAVigation Satellite System (GLONASS), and the forthcoming Galileo transmitters, in medium Earth orbit, provide signals with well-controlled quality, constant level, and guarantee that at least four space-based sources of opportunity are visible at any position on the earth at any time [22, 23].
- Low Earth orbit based transmitters for satellite personal communication [24].
- Geostationary/geosynchronous telecommunication satellites for audio and video broadcast, data exchange, and personal communications offer the coverage of wide continental/sea surface areas with significant Equivalent Isotropic Radiated Power (EIRP) levels at medium latitudes at the earth's surface [25, 26].

The low power signal reflected from the target is collected by the main PBR receiver (typically known as the surveillance channel) using a directive antenna steered toward the surveillance area. Since the transmitted signal is not known at the receiver, an auxiliary PBR receiver (typically known as the reference channel) is usually connected to an additional directive antenna steered toward the transmitter. The typical basic PBR processing scheme is depicted in Figure 17-1.

The signal collected at the reference channel is first used to remove undesired contributions that have been received, along with the moving target echo, on the surveillance channel. Such disturbances result from the fraction of the direct signal coming from the transmitter and received by the sidelobes or backlobes of the surveillance antenna, as well as the strong clutter/multipath echoes. This problem can have major relevance when the direct signal contribution is strong and affected by multiple reflections on the surface or on discrete scatterers. Various approaches have been proposed to cope with this problem. After the cancellation stage, the detection process is based on the evaluation of the bistatic two-dimensional (range-velocity) cross-correlation function (2D-CCF) between the surveillance and the reference signal. A constant false alarm rate (CFAR) threshold can be then applied on the obtained map to automatically detect the potential targets according to a specific CFAR detection scheme.

While it is not especially challenging to evaluate 2D-CCF from a signal processing point of view, it can be in terms of the computational load that it requires for real-time application. The long integration times (order of magnitude of seconds) required to extract the low-level target signal from the disturbance implies a large number of complex multiplications to be performed. Moreover, long coherent integration times of this type provide

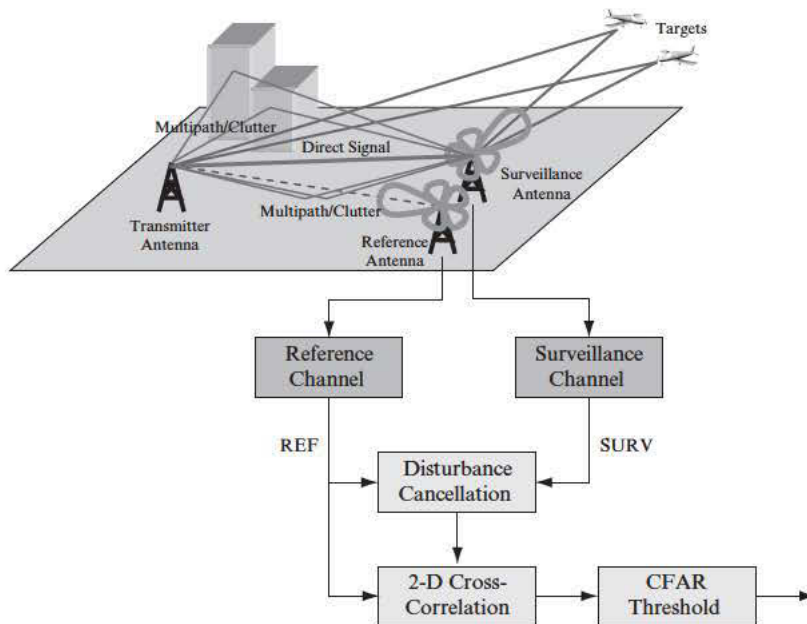


FIGURE 17-1 ■
Basic PBR
processing scheme.

a high Doppler frequency resolution; thus, a high number of Doppler filters is required. Finally, the computational load is also proportional to the bandwidth of the specific waveform of opportunity used in the system because the bandwidth sets the minimum required sampling frequency. Therefore, evaluation of the 2D-CCF for wide bandwidth waveforms, like WiFi and WiMAX, has a very high computational cost. By observing that the expected target echo belongs to a limited area of the bistatic range-Doppler plane, the computational load can be lowered by appropriately selecting the filter implementation approach that best matches the region of the bistatic range-Doppler plane of interest and by considering approximate filtering schemes. A review of these techniques is presented in Section 17.2.

The disturbance cancellation technique is of particular importance when it is difficult to avoid a significant signal power arriving directly from the transmitter of opportunity into the sidelobes of the surveillance antenna. For a long range target, the power ratio of direct signal received through the sidelobes and desired signal can get close to 100 dB so that the cancellation technique has a significant challenge. This is especially the case of passive radar operating at relatively low frequency, where it is very difficult to reduce the receiving antenna sidelobes to a very low level. This is clearly experienced with FM-based PBR. A review of signal processing techniques for disturbance cancellation in PBR radar is presented in Section 17.3, showing approaches with different complexity and effectiveness. We also show solutions potentially yielding remarkable cancellation performance and acceptable computational load.

While the basic detection scheme is fully defined by the described cancellation and integration techniques, the effectiveness of the passive radar operation can be largely enhanced by considering an advanced PBR processing scheme as the one sketched in Figure 17-2.

First we notice that both 2D-CCF and interference cancellation assume a perfect knowledge of the transmitted waveform. In contrast, in practice the transmitted waveform is estimated by means of the reference channel, and that is affected by all transmitter, propagation, and receiver modifications. Therefore, it might be necessary to perform some

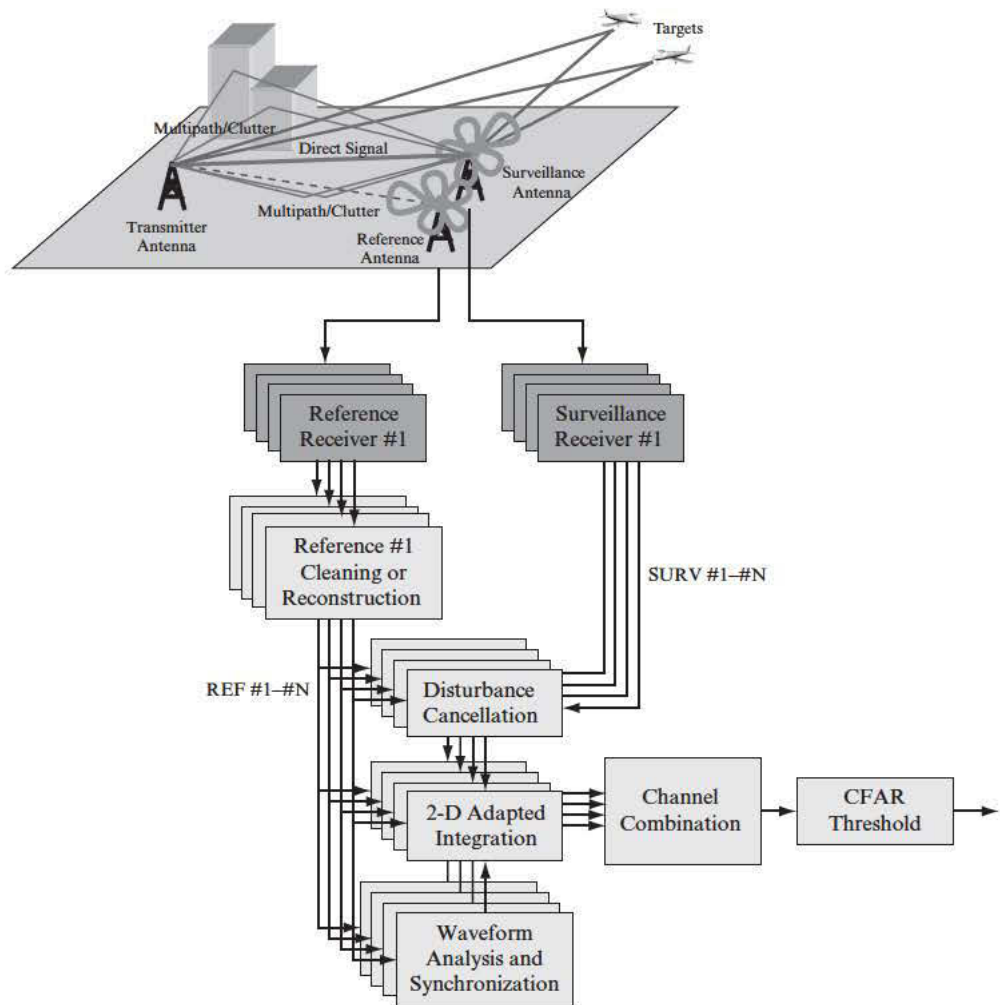


FIGURE 17-2 ■ Advanced PBR processing scheme.

transmitter-specific conditioning of the received signal before cancellation and 2D-CCF evaluation, and this depends on the specific waveform of opportunity exploited. These may include high-quality analog bandpass filtering, channel equalization, removal of unwanted structures in digital signals, or complete reconstruction from the received digital signal. Such conditioning approaches are mainly intended for the reference signal to improve:

- a. Its quality (with a benefit for the cancellation step). This aspect is addressed in Section 17.4 that analyzes the effect of multipath contributions in the reference signal and shows that these contributions represent the main cause of degradation of the reference signal quality. Proper techniques are introduced to remove such undesired contributions from the reference signal for PBR systems exploiting both analog and digital transmissions; in the latter case, the possibility to demodulate them and reconstruct their multipath-free version is explored.
- b. The resulting ambiguity function (AF). PBR operation inherently implies that the transmitted waveform is not within the control of the radar designer. This contrasts

to the usual case of conventional radar systems, where the transmitted waveform is carefully designed to provide an AF with appropriate properties (e.g., narrow peak in both range and Doppler and low sidelobes). This aspect is investigated in Section 17.5 with particular reference to digital transmissions, which are usually characterized by a number of undesired peaks or high sidelobes in the corresponding 2D-CCF and have a severe masking effect on small targets. Therefore, proper techniques are introduced to cope with this undesired effect in PBR systems based on digital transmissions exploiting different modulation schemes.

A significant performance improvement for a PBR system is expected when the receiver is equipped with multiple receiving/processing channels as sketched in Figure 17-2. These might be exploited to simultaneously collect and process multiple signals received at different frequencies, different polarizations, or different antennas. Section 17.6 shows some examples of how to exploit the aforementioned multichannel information and the available increase of performance. In particular, the use of multiple-frequency channels is shown to provide an increased robustness for the detection, and the joint use of multiple polarimetric channels is exploited to better remove the effect of undesired interferences from other transmitters. Moreover, the use of the multiple channels corresponding to antenna arrays is shown to offer the potentialities to steer the receiving beams and also to use adaptive beamforming to optimize the PBR target detection capability.

17.1.1 Organization

This chapter is devoted to the presentation of the advanced signal processing techniques that are applied at the passive bistatic radar (PBR) receiver to perform optimized detection of air and surface moving targets.

The Introduction above gives an overview of the state-of-the-art on passive radar by providing a list of the many different transmitters that are available for telecommunication, radio-navigation, and remote sensing applications and have been used in the past as sources of opportunity for PBR. The PBR principle is described and the main steps of the PBR processing scheme are briefly summarized.

Section 17.2 reports a review of the techniques for the evaluation of the bistatic two-dimensional (range-velocity) cross-correlation function (2D-CCF). Specifically the challenges related to the required computational load are illustrated and both optimum and sub-optimum techniques are described to be exploited for real time operation.

The PBR limitations due to disturbance contributions (both direct signal breakthrough and clutter echoes) are investigated in Section 17.3, where a review of the signal processing techniques for disturbance cancellation is presented, showing approaches with different complexity and effectiveness.

In Section 17.4 the effect of multipath contributions in the reference signal is analyzed and proper techniques are introduced to remove such undesired contributions in PBR systems exploiting both analog and digital transmissions.

The problem of the ambiguity function (AF) control in PBR is investigated in Section 17.5 and proper techniques are introduced for PBR systems based on digital transmissions exploiting different modulation schemes.

Section 17.6 gives a taste of the potentialities achievable for PBR operation when using multichannel receiving systems, with specific reference to PBR sensors characterized by either multifrequency, polarimetric, or adaptive antenna array capability.

Finally few conclusions are drawn in Section 17.7 on the potentialities and further processing requirements for the future passive radar sensors.

17.1.2 Key Points

Important concepts developed throughout the chapter are summarized as follows:

- PBR exploits existing illuminators of opportunity which results in the exciting possibility of low cost surveillance, covert operation, and reduced electromagnetic pollution.
- PBR practical feasibility for surveillance purposes has been well established. Nevertheless, there are still many challenges to be solved involving both technology development and processing techniques.
- A dedicated and separate receiver channel is required to collect the transmitted signal. This is subsequently used as the reference signal and is correlated with the surveillance signal for target echo detection.
- The evaluation of the bistatic 2D-CCF represents one of the most costly processing steps in terms of computational burden due to the long coherent integration times. For real time operation, cost-effective solutions can be adopted based on both optimum and sub-optimum techniques.
- Disturbance cancellation is a required step in the typical PBR processing scheme since it allows the removal of undesired contributions received on the surveillance channel along with the moving target echo (i.e. direct signal from the transmitter and strong clutter/multipath echoes).
- The reference signal is usually assumed to be a high-quality copy of the transmitted signal. In practice, some transmitter-specific conditioning of the reference signal may be performed to improve its quality and hence the PBR performance.
- PBR operation inherently implies that the transmitted waveform is not within the control of the radar designer. However, in specific cases, proper filters can be used to control the sidelobes of the resulting AF.
- With the aim to increase the performance and the reliability of PBR systems, advanced multichannel processing techniques can be applied. These might exploit parallel receiving/processing channels that differ by frequency, polarization, time, and space.

17.1.3 Notation

The following lists many of the variable names found within this chapter:

$s_{ref}(t), s_{ref}[n]$ = reference signal complex envelope in continuous and discrete time notations;

$s_{surv}(t), s_{surv}[n]$ = surveillance signal complex envelope in continuous and discrete time notations;

$\chi(\tau, f_D), \chi[l, m]$ = 2– dimensional cross-correlation function (2D-CCF) in continuous and discrete time notations;

τ = bistatic time difference of arrival (s);

f_D = bistatic Doppler (Hz);

f_s = sampling frequency (Hz);

R_B = relative bistatic range, i.e. the difference between the 2-way path length and the transmitter-receiver baseline (m);

c = velocity of propagation (speed of light, m/s)
 v_B = bistatic velocity, i.e. the rate of change of the bistatic range (m/s);
 λ = center wavelength (m);
 T_{int} = coherent processing interval (s);
 N = number of integrated samples;
 N_τ = number of time bins in the Range-Velocity map;
 N_f = number of Doppler bins in the Range-Velocity map;
 $L[l, m]$ = SNR loss at the l -th time bin and m -th Doppler bin for sub-optimum 2D-CCF evaluation;
 n_B = number of batches used by the Batches algorithm for 2D-CCF evaluation;
 N_B = number of samples for each batch used by the Batches algorithm;
 n_C = number of frequency bands (channels) used by the Channelization technique for 2D-CCF evaluation;
 N_C = number of batches used by the Channelization technique;
 $d(t)$ = complex envelope of the direct signal (a delayed replica of the transmitted signal);
 A_{surv} = complex amplitude of the direct signal received at the surveillance antenna;
 N_T = number of targets;
 a_m = complex amplitude of the echo from the m -th target;
 τ_{Tm} = delay (with respect to the direct signal) of the echo from the m -th target (s);
 f_{Dm} = Doppler frequency of the echo from the m -th target (Hz);
 N_S = number of ground scatterers;
 c_i = complex amplitude of the echo from the i -th stationary ground scatterer;
 τ_{ci} = delay (with respect to the direct signal) of the echo from the i -th ground scatterer (s);
 $n_{surv}(t)$ = thermal noise contribution at the surveillance channel;
 $\mathbf{s}_{surv}, \mathbf{s}_{ref}$ = vector of complex samples collected at the surveillance channel and at the reference channel;
 K = number of taps (degrees of freedom) used by the ECA algorithm;
 α = weight vector for the ECA algorithm;
 \mathbf{s}_{ECA} = vector of surveillance signal samples after ECA processing;
 b = number of batches used by the ECA-B algorithm;
 T_b = temporal duration of each batch used by the ECA-B algorithm (s);
 N_b = number of samples for each batch used by the ECA-B algorithm;
 P_{fa} = Probability of False Alarm;
 A_{ref} = complex amplitude of the direct signal received at the reference antenna;
 A_m = complex amplitude of the m -th multipath replica;
 τ_m = delay (with respect to the direct signal) of the of the m -th multipath replica (s);
 ϕ_m = phase shift between direct signal and m -th multipath replica;
 θ_m = direction of arrival (DOA) of the m -th multipath replica;
 $n_{ref}(t)$ = thermal noise contribution at the reference channel;
 M = number of array elements (receiving channels);
 L = number of taps (temporal degrees of freedom) used by the CMA algorithm;
 $w_{m,l}[n]$ = adaptive weight at the l -th tap on the m -th channel at the n -th time instant for the CMA algorithm;
 μ = step-size parameter of the CMA algorithm;
 N_P = number of pulses within the CPI for a pulsed transmission;
 $\chi^{(m)}(\tau), \chi^{(m)}[l]$ = temporal Cross-Correlation Function (CCF) at the m -th pulse in continuous and discrete time notations;

\mathbf{c} = 11-chip Barker sequence;
 T_{sym} = symbol duration (s);
 T_{chip} = chip duration (s);
 $\alpha_B[n]$ = n -th sample of the Auto-Correlation Function (ACF) of the 11-chip Barker sequence;
 $h_{BWN}[n]$ = n -th weight of the Barker Weighting Network (BWN);
 L_{BWN} = filter length for the Barker Weighting Network;
 $\alpha_{BWN}[n]$ = n -th sample of the ACF of the 11-chip Barker sequence after the BWN application;
 K_S = number of range sidelobes to be removed by the ASRF algorithm;
 $\chi_{ASRF}^{(m)}(\tau)$ = temporal Cross-Correlation Function (CCF) at the m -th pulse after the ASRF application;
 N_{FFT} = FFT size in OFDM transmissions;
 Δf = subcarriers spacing in OFDM transmissions (Hz);
 T_U = useful symbol duration in OFDM transmissions (s);
 T_G = guard interval in OFDM transmissions (s);
 T_S = total symbol duration in OFDM transmissions (s);
 N_{ch} = number of integrated channels in the multi-frequency approach;
 f_n = center frequency for the n -th frequency channel (Hz);
 Δv_n = velocity resolution for the n -th frequency channel (m/s);
 P = number of secondary data used by the CFAR detection algorithm;
 G_{CA} = threshold used by the CA-CFAR detection algorithm;
 \mathbf{s}_θ = target steering vector;
 $g_i(\theta)$ = pattern of the i -th element of the array;
 R = radius of circular arrays (m);
 \mathbf{M} = disturbance covariance matrix;
 $\mathbf{w}(\theta_0)$ = spatial weight vector for the array beam steered toward angle θ_0 .

17.1.4 Acronyms

Commonly used acronyms used in this chapter include:

2D-CCF	two-dimensional cross-correlation function
ACF	auto-correlation function
AF	ambiguity function
AP	access point
ASRF	additional sidelobes reduction filter
ATC	air traffic control
BWN	weighting network for Barker sidelobes reduction
CA	clutter attenuation
CA-CFAR	cell-averaging CFAR
CCF	cross-correlation function
CFAR	constant false alarm rate
CIC	cascaded integrator-comb
CMA	constant modulus algorithm
CPI	coherent processing interval
DAB	digital audio broadcasting
DFT	discrete Fourier transform
DL	down-link

DMR	direct signal-to-multipath ratio
DNR	direct signal-to-noise ratio
DOA	direction of arrival
DOF	degrees of freedom
DSSS	direct sequence spread spectrum
DVB-T	digital video broadcasting - terrestrial
ECA	extensive cancellation algorithm
ECA-B	ECA batches
ECA-B&S	ECA batches & stages
FFT	fast Fourier transform
FIR	finite impulse response
FM	frequency modulation
FUSC	fully used sub-channelization
GAL	gradient adaptive lattice
HF	high frequency
IDFT	inverse discrete Fourier transform
IFFT	inverse fast Fourier transform
MF	multiple-frequency
MFC	multiple-frequency configuration
MTI	moving target indication
NLMS	normalised least mean squares
OFDM	orthogonal frequency-division multiplexing
PBR	passive bistatic radar
PNFR	peak-to-noise floor ratio
PSD	power spectral density
PSK	phase-shift keying
PSLR	peak-to-sidelobe ratio
PUSC	partially used sub-channelization
Rx	receiver
S-CMA	space constant modulus algorithm
SFC	single frequency channel
SNR	signal-to-noise ratio
ST-CMA	space-time constant modulus algorithm
T-CMA	time constant modulus algorithm
TPS	transmission parameter signalling
Tx	transmitter
UHF	ultra-high frequency
UL	up-link
WiMAX	worldwide interoperability for microwave access

17.2 | EVALUATION OF THE 2D-CCF FOR THE PASSIVE RADAR COHERENT INTEGRATION

The evaluation of the bistatic range-velocity 2D-CCF is the key step in the PBR processing chain. It corresponds to the implementation of a bank of matched filters, each one tuned to a specific target bistatic velocity. Typically the values of velocity are chosen with a separation approximately equal to the bistatic velocity resolution. Therefore, the set

of filters covers all the possible target velocities. The filter where the target is detected provides the estimates of the bistatic Doppler shift of each target echo. Similarly, the estimate of the bistatic range is given by the sample along range where it is detected. Assuming that the signal $s_{ref}(t)$ collected at the reference antenna is a perfect copy (or at least good enough as discussed in a later section) of the transmitted signal, the 2D-CCF for a PBR is evaluated as

$$\chi(\tau, f_D) = \int_{-\infty}^{\infty} s_{surv}(t) s_{ref}^*(t - \tau) e^{-j2\pi f_D t} dt \quad (17.1)$$

where

$s_{surv}(t)$ = the complex envelope of the signal collected at the surveillance antenna

$\tau = R_B/c$ = the bistatic time difference of arrival of interest, where R_B is the relative bistatic range, or the difference between the two-way path length and the baseline between the transmitter (Tx) and the receiver (Rx)

$f_D = v_B/\lambda$ = the Doppler shift of interest, v_B , being the bistatic velocity (i.e., the rate of change of Tx – target – Rx range)

The integral is limited in practice to the coherent processing interval (CPI) T_{int} .

Assuming that the signals are sampled at frequency f_s , equation (17.1) can be easily expressed in discrete time notation as

$$\chi[l, m] = \sum_{n=0}^{N-1} s_{surv}[n] \cdot s_{ref}^*[n - l] \cdot e^{-j2\pi \frac{mn}{N}} \quad (17.2)$$

where

$N = \lfloor T_{int} f_s \rfloor$ = the number of integrated samples

l = the time bin corresponding to time delay $\tau = l/f_s$

m = the Doppler bin corresponding to Doppler shift $f_D = mf_s/N$

The evaluation of the 2D-CCF for a PBR represents one of the most costly operations in terms of computational burden [4,27,28]. In fact, the exploited waveform of opportunity typically has a low power level for radar purposes, so a very long integration time is usually required to obtain an acceptable signal-to-noise ratio (SNR). Moreover, large 2-D maps might be required depending on the desired surveillance region extent in both range ($0, R_{B\max}$) and Doppler dimensions ($-|v_B|_{\max}, |v_B|_{\max}$), where $R_{B\max}$ and $|v_B|_{\max}$ are the maximum relative bistatic range and the maximum bistatic velocity of interest, respectively. This implies that a huge amount of data has to be managed and a large number of complex operations has to be performed that might require very fast hardware for real-time processing. As a reference, the direct evaluation of (17.2) over N_τ time bins and N_f Doppler bins requires $N(N_\tau + 1)N_f$ complex multiplications and $(N - 1)N_\tau N_f$ complex additions, where, assuming that the 2D-CCF is not oversampled, we have

$$N_\tau = \left\lceil \frac{R_{B\max}}{c} f_s \right\rceil \quad \text{and} \quad N_f = \left\lceil \frac{2|v_B|_{\max}}{\lambda} \frac{N}{f_s} \right\rceil \quad (17.3)$$

17.2.1 Efficient Implementations of the 2D-CCF Based on the Fast Fourier Transform

Two alternative efficient implementations can be obtained by rearranging the factors in (17.2) according to different criteria and exploiting the well-known fast Fourier transform (FFT) algorithm to evaluate the discrete Fourier transform (DFT).

Correlation FFT

This algorithm is based on the observation that, at the m -th Doppler bin, the samples along time of the 2D-CCF correspond to the samples of the cross-correlation between the reference signal and a Doppler shifted version of the surveillance signal $s_m[n] = s_{surv}[n] \exp(-j2\pi mn/N)$:

$$\chi[l, m] = \sum_{n=0}^{N-1} s_m[n] \cdot s_{ref}^*[n-l] = C_m[l] \quad (17.4)$$

A saving in computation is obtained by evaluating such cross-correlation in the frequency domain as

$$C_m[l] = IDFT\{DFT\{s_m[n]\}(DFT\{s_{ref}[n]\})^*\} = IDFT\{S_m[k]S_{ref}^*[k]\} \quad (17.5)$$

For each one of the N_f Doppler filters, an inverse DFT (IDFT) is evaluated and $N - N_\tau$ output samples are discarded. Figure 17-3a is a sketch of the resulting algorithm. We notice that, for a given m , the DFT $S_m[k]$ of $s_m[n]$ can be simply obtained as a circular shift of the samples of $S_{surv}[k] = DFT\{s_{surv}[n]\}$. Consequently, the DFTs of the reference and surveillance signals can be evaluated just once and, at each subsequent iteration, only

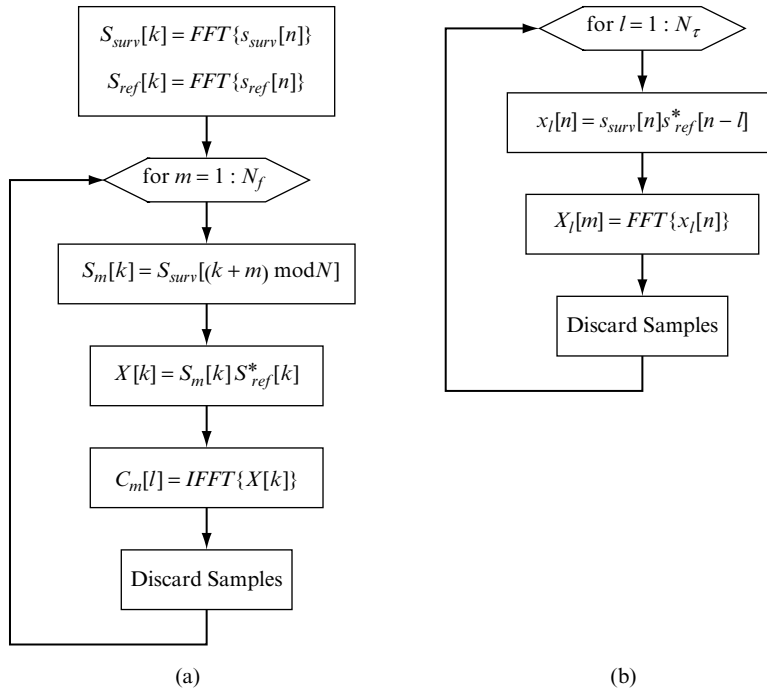


FIGURE 17-3 ■ Sketch of efficient algorithms for the 2D-CCF evaluation. (a) Correlation FFT. (b) Direct FFT.

TABLE 17-1 ■ Computational Loads for Optimum and Suboptimum Algorithms for 2D-CCF Evaluation

Algorithm	Complex Multiplications	Complex Additions
Correlation FFT	$2N \log_2(N) + N_f[N + N \log_2(N)]$	$2N \log_2(N) + N_f[N \log_2(N)]$
Direct FFT	$N_\tau[N + N \log_2(N)]$	$N_\tau[N \log_2(N)]$
Batches algorithm	$N_\tau[N + n_B \log_2(n_B)]$	$N_\tau[N + n_B \log_2(n_B)]$
Decimation using CIC filters	$N_\tau[N + 5N_D + N_D \log_2(N_D)]$	$N_\tau[N + N_D + N_D \log_2(N_D)]$
Channelization technique	$2N \log_2(n_C) + N + N \log_2(N)$	$2N \log_2(n_C) + N \log_2(N)$

the N complex multiplications and a single inverse FFT (IFFT) must be performed. The resulting computational load is reported in Table 17-1. This algorithm can be implemented in parallel for the N_f Doppler bins of interest.

Direct FFT

This algorithm is based on the observation that, at the l -th time bin, the samples along Doppler frequency of the 2D-CCF correspond to the samples of the DFT of the sequence $x_l[n] = s_{surv}[n]s_{ref}^*[n-l]$ obtained as the product of the surveillance signal and the delayed conjugated reference signal, [4]:

$$\chi[l, m] = X_l[m] = DFT\{x_l[n]\} \quad (17.6)$$

Similar to the previous case, for each one of the N_τ bistatic delays, the DFT is evaluated, and $N - N_f$ output samples are discarded. The resulting algorithm is sketched in Figure 17-3b. Since its iterations involve the range dimension, this algorithm can be parallelized over the range bins and can be limited to the N_τ range bins of interest. Therefore, at each iteration only N complex multiplications and a single FFT must be performed (in contrast to the correlation FFT algorithm it does not require the preliminary evaluation of the signals DFT). The resulting computational load is reported in Table 17-1 for comparison.

As is apparent from Table 17-1, the computational load for both algorithms increases with the number of integrated samples as $N \log_2(N)$. However, the final cost of the correlation FFT is essentially determined by the number N_f of considered Doppler bins, while the cost of the direct FFT is essentially determined by the number N_τ of range bins included in the 2-D map. Thus, the algorithm with the lowest number of operations depends on the extent of the 2D-CCF over the range and Doppler dimensions required for the specific application: if $N_\tau < N_f$ the direct FFT algorithm requires less computation than the correlation FFT.

As an example, Table 17-2 reports the numerical evaluation of the computational load of both the correlation FFT and the direct FFT for an FM-based PBR (the main parameters of the considered study case are also listed in the table). It is easily verified that, as $N_\tau \cong 3/5 \cdot N_f$, the direct FFT algorithm requires almost half the computational load of the correlation FFT. Notice that this is not always the case. For example, let us consider a PBR based on DVB-T transmissions aimed at the surveillance of the same area described in Table 17-2 in both range and velocity dimensions. In this case, a signal bandwidth of 8 MHz might be assumed at a carrier frequency of 500 MHz (ultra high frequency [UHF] band). These requirements result in a 2D-CCF consisting of $N_\tau = 6400$ range bins and $N_f = 256$ Doppler bins (assuming that the integration time has been properly modified to yield the same velocity resolution). The benefits of the correlation

TABLE 17-2 ■ Example of Computational Load Comparison for Optimum and Suboptimum Algorithms for 2D-CCF Evaluation for a FM-Based PBR

Study case description	Algorithm	Parameter selection	Complex Multiplications	Complex Additions	Gain Loss (dB)
FM-based PBR Carrier frequency: 100 MHz $T_{int} = 1$ sec $N = 2 \cdot 10^5$ $R_{B \max} = 240$ km $N_\tau = 160$ $ v_B _{\max} = 384$ m/s $N_f = 256$	Correlation FFT	–	$9.6 \cdot 10^8$	$9.1 \cdot 10^8$	–
	Direct FFT	–	$5.9 \cdot 10^8$	$5.6 \cdot 10^8$	–
	Batches algorithm	$n_B = 500$	$3.3 \cdot 10^7$	$3.3 \cdot 10^7$	< 1 dB
	Decimation using CIC filters	$D = 128$	$3.6 \cdot 10^7$	$3.5 \cdot 10^7$	Negligible
	Channelization technique	$n_C = 500$	$7.3 \cdot 10^6$	$7.1 \cdot 10^6$	< 4 dB

FFT with respect to the direct FFT are remarkable in this case. A similar conclusion might be drawn when assuming that the same integration time considered for the FM-based PBR in Table 17-2 ($T_{int} = 1$ sec) is maintained, thus obtaining a number of Doppler bins $N_f = 1280$. However, with wider bandwidth signals the integration time might need to be reduced to account for the expected target range migration.

17.2.2 Suboptimum Implementations of the 2D-CCF

The computations required by the 2D-CCF can be further reduced by resorting to suboptimum algorithms, if small degradations can be accepted in term of SNR. The required cost for both the correlation FFT and the direct FFT optimum algorithms is strongly affected by the processing load required by the FFT of the long input sequences. Furthermore, only a very small portion of the output FFTs (namely, N_τ out of N , with $N_\tau \ll N$, or N_f of N , with $N_f \ll N$) is required in the final 2-D map, and most of the obtained samples are discarded. Therefore, the computational load can be reduced even further by avoiding useless computations and by approximating the phase compensations to be performed by appropriately decimating or subdividing in channels the signals according to the following algorithms.

Batches Algorithm

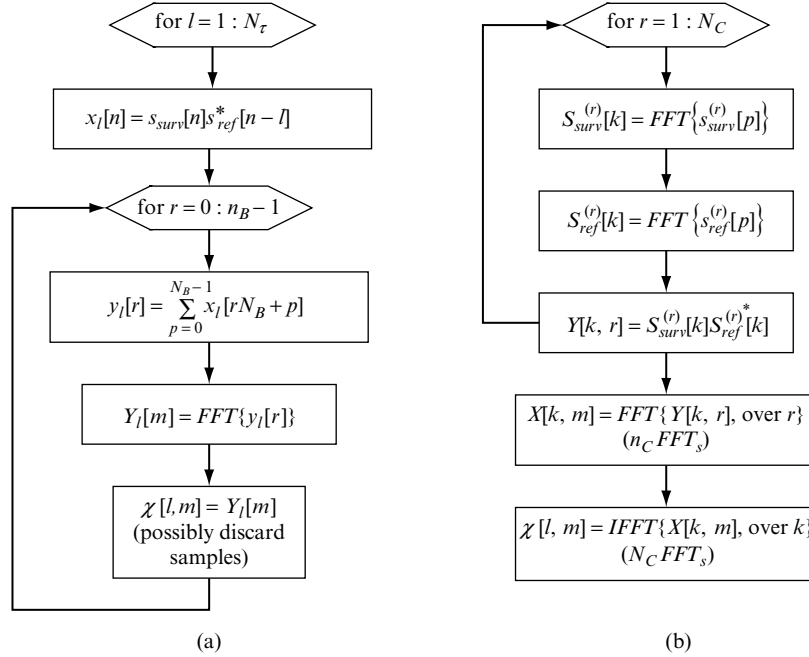
This suboptimum algorithm is obtained by subdividing the CPI into n_B batches (sub-CPIs) of N_B samples each. The Doppler frequency compensation inside each short sub-CPI is neglected but maintains the phase compensation from batch to batch and thus reduces the overall computation complexity with limited loss in signal processing gain. Specifically, equation (17.2) is rewritten as

$$\chi[l, m] = \sum_{r=0}^{n_B-1} e^{-j2\pi \frac{mr}{N}} \sum_{p=0}^{N_B-1} s_{surv}[rN_B + p] \cdot s_{ref}^*[rN_B + p - l] \cdot e^{-j2\pi \frac{mp}{N}} \quad (17.7)$$

and neglecting the complex exponential term within the inner summation we obtain

$$\chi[l, m] \cong \sum_{r=0}^{n_B-1} e^{-j2\pi \frac{mr}{N}} \sum_{p=0}^{N_B-1} s_{surv}[rN_B + p] \cdot s_{ref}^*[rN_B + p - l] = \sum_{r=0}^{n_B-1} e^{-j2\pi \frac{mr}{n_B}} y_l[r] \quad (17.8)$$

FIGURE 17-4 ■ Sketch of suboptimum algorithms for 2D-CCF evaluation. (a) Batches algorithm. (b) Channelization technique.



that represents the DFT of the decimated sequence $y_l[r]$ ($r = 0, \dots, n_B - 1$), obtained by summing the product sequence $x_l[n] = s_{\text{surv}}[n] \cdot s_{\text{ref}}^*[n - l]$ for each batch. The resulting algorithm is sketched in Figure 17-4a, and its computational load is reported in Table 17-1 as a function of the number of batches assuming that $n_B N_B = N$.

The number of batches should be selected so that $n_B \geq N_f$, and, in particular, no sample is discarded at the FFT output if equality holds. Within this limit, a smaller n_B provides a lower computational load but also yields a higher integration loss for fast-moving targets whose Doppler shift is not properly compensated [28]. In fact, for a constant modulus signal the SNR loss for a target at range bin l and Doppler bin m can be written

$$L[l, m] = -20 \log_{10} \left\{ \frac{n_B \sin(\pi m / n_B)}{N \sin(\pi m / N)} \right\} \quad (17.9)$$

As is apparent, it depends only on the considered Doppler bin and reaches its maximum at the highest Doppler value considered in the 2D-CCF, while no loss is experienced at zero Doppler.

As an example, Figure 17-5a shows maximum SNR loss (for the highest Doppler frequency) and computational load (number of complex multiplications) as a function of the number of batches n_B for the study case described in Table 17-2. When operating with n_B between 500 and 2500, the batches algorithm yields a computational load about 17 times smaller than the direct FFT, with an SNR loss below 1 dB in the worst case. This is confirmed by Table 17-2 reporting the results of the batches algorithm for $n_B = 500$ providing a reduction in computation by a factor 18.2. Notice that the same 1 dB loss would be experienced by the direct FFT operating with a number of integrated samples reduced by a factor of 1.2. However, in this case for the same SNR loss, the computational load would be reduced by only a factor of 1.2.

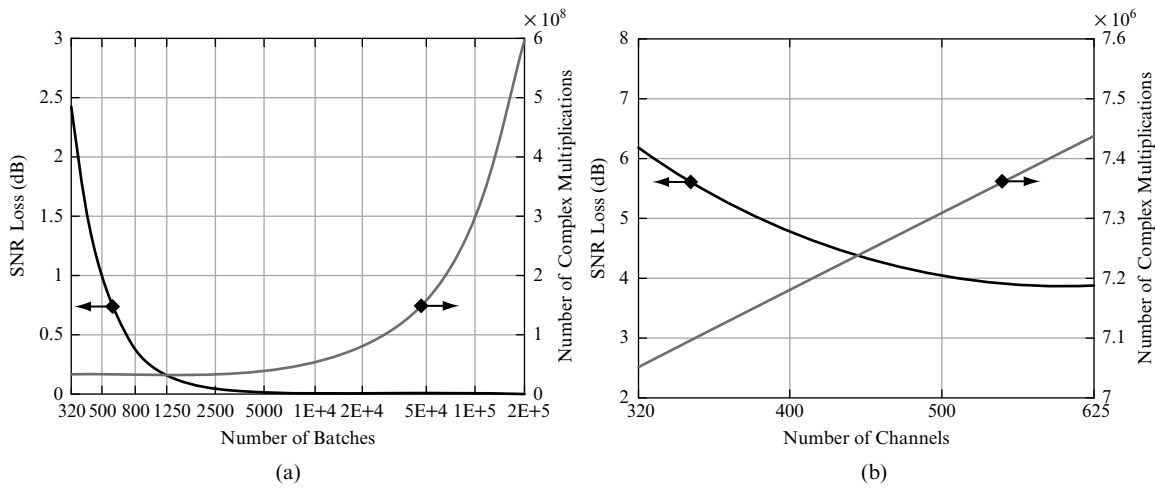


FIGURE 17-5 ■ SNR loss and computational load versus selected parameter for suboptimum algorithms for 2D-CCF evaluation for a FM-based PBR (study case described in Table 17-2). (a) Batches algorithm. (b) Channelization technique.

As previously mentioned, the desired cost reduction has been obtained by summing the product sequence $x_l[n]$ over each batch of length N_B . This is equivalent to applying a low-pass filter with rectangular impulse response to the signal $x_l[n]$. The shape of the 2D-CCF with frequency at the l -th range bin and the SNR loss are direct consequences of the typical $\sin(xN_B)/\sin(x)$ shape of the transfer function of this specific filter. Following these considerations, different low-pass filters can be used to have different shapes for the SNR loss versus frequency bin while still providing the possibility to apply decimation to the sequence $x_l[n]$. The rectangular filter does not require complex multiplications, whereas the generic filter might, thus trading the lower SNR loss with computations. Obviously, many possible solutions can be used; in particular, using a cascaded integrator-comb (CIC) filter to perform the low-pass operation does not imply any increase in required computations while providing lower losses. To improve the shape of the CIC filter, it is customary to use in cascade with it a finite impulse response (FIR) filter with a small number of taps (thus providing a limited increase in required computations) that is able to control the shape of the CIC transfer function (and therefore reduce the loss with target Doppler frequency). As an example, a possible approach is presented in [4] where a CIC filter operating with a decimation factor $D = 128$ ($N_D = N/D$) is considered together with a fifth-order low-pass (symmetric) FIR filter. The resulting computational load is reported in Table 17-1 and evaluated in Table 17-2 for our study case using a filter with cut-off frequency equal to 300 Hz. As for the direct FFT, the 2D-CCF evaluation might be parallelized over the range bins and limited to some range bins of interest.

Channelization Technique

This suboptimum algorithm operates by splitting the signals in the frequency domain into n_C frequency bands (channels) that are separately processed. The results can be coherently recombined to build the final map with limited gain loss. It can be easily obtained as a

further approximation of the batches algorithm described already. To show its derivation, we first rewrite equation (17.2) and then replace the cross-correlation between the r -th portions of length n_C of the surveillance and the reference signals, with its approximation evaluated in the frequency domain:

$$\begin{aligned}
 \chi[l, m] &\cong \sum_{r=0}^{N_C-1} e^{-j2\pi \frac{mr}{N_C}} \sum_{p=0}^{n_C-1} s_{\text{surv}}[rn_C + p] \cdot s_{\text{ref}}^*[rn_C + p - l] \\
 &\cong \sum_{r=0}^{N_C-1} e^{-j2\pi \frac{mr}{N_C}} \frac{1}{n_C} \sum_{k=0}^{n_C-1} S_{\text{surv}}^{(r)}[k] S_{\text{ref}}^{(r)*}[k] e^{+j2\pi \frac{kl}{n_C}} \\
 &= \frac{1}{n_C} \sum_{k=0}^{n_C-1} e^{+j2\pi \frac{kl}{n_C}} \sum_{r=0}^{N_C-1} Y[k, r] e^{-j2\pi \frac{mr}{N_C}} \quad (17.10)
 \end{aligned}$$

Notice that by neglecting the border effects (additional approximation with respect to the batches algorithm), the r -th subsequences are cyclically extended so that their cross-correlation can be evaluated in the frequency domain, as the IDFT of their individual DFTs ($S_{\text{surv}}^{(r)}[k]$ and $S_{\text{ref}}^{(r)*}[k]$ in the second line). By rearranging the terms of the summations, the third line of equation (17.10) is easily obtained.

The resulting algorithm (sketched in Figure 17-4b) can be subdivided into four steps:

1. Evaluate the DFTs of surveillance and reference signals over the N_C batches of n_C samples each.
2. Compute the product of the DFTs outputs for each batch and arrange the results in the 2-D sequence $Y[k, r] = S_{\text{surv}}^{(r)}[k] S_{\text{ref}}^{(r)*}[k] (k = 0, \dots, n_C - 1; r = 0, \dots, N_C - 1)$.
3. Evaluate the DFT $X[k, m]$ of $Y[k, r]$ over the index r for each k (n_C IDFTs); this step represents the parallel processing of the n_C different frequency channels in which the signals have been split.
4. Evaluate the IDFT of $X[k, m]$ over the index k for each m (N_C DFTs) to obtain the range samples of the 2D-CCF at the m -th Doppler bin.

Notice that evaluating the DFT of short subsequences (Step 1) separates each sequence into a number of frequency channels (channelization) that are independently processed (Steps 2 and 3) and then recombined to provide the final result (Step 4).

The computational load for the channelization technique is reported in Table 17-1 as a function of the number n_C of channels assuming $n_C N_C = N$. In this case, the number of frequency channels is equal to the batches length; thus, the algorithm parameters must be selected so that $N_C \geq N_f$. Moreover, since the range samples of the 2D-CCF are obtained from the last DFT stage performed over sequences of n_C samples, assuming that only positive values of the differential bistatic range R_b are included in the 2D-CCF map, $n_C \geq 2N_\tau$ should be satisfied. Reducing n_C within these limits allows a corresponding reduction in the computational load.

The number of channels should be carefully selected since it yields a trade-off between the two adopted approximations of equation (17.10) that determine the final SNR loss. Increasing n_C yields a higher integration loss for fast-moving targets whose Doppler shift is not properly compensated (as for the batches algorithm) [28]. However, in this case, reducing n_C enhances the border effects in the evaluation of the batches cross-correlations yielding additional loss for targets at higher bistatic ranges. Assuming a constant modulus

signal, the SNR loss for a target at range bin l and Doppler bin, m can be written as

$$L[l, m] = -20 \log_{10} \left\{ \frac{1}{n_C} \frac{\sin [\pi m (n_C - l)/N]}{\sin (\pi m / N)} \right\} \quad (17.11)$$

For $l = 0$, the same loss of the batches algorithm given in equation (17.9) is obtained at the m -th Doppler bin, but the SNR loss for a given n_C increases at higher range bins. Thus, the integration loss is maximized at the highest Doppler and range values considered in the 2D-CCF.

As an example, Figure 17-5b shows maximum SNR loss and computational load (number of complex multiplications) as a function of the number n_C of channels for the study case described in Table 17-2. For any number of channels within the allowed range, the channelization technique yields a computational load reduction of two orders of magnitude with respect to the optimum techniques but might yield an SNR loss of 6 dB if n_C is the minimum allowable value. Table 17-2 reports the results of the channelization technique operating with $n_C = 500$, which limits the SNR loss to 4 dB in the worst location within the 2D-CCF but requires the smallest number of complex operations.

SNR loss evaluation gives only limited insight into the surveillance capability of the resulting PBR system. However, the sources of opportunity usually provide CW waveforms, and, for the real-time operation, only a fraction of the available signals is typically used for signal detection due to the limited computational resources available. Under these conditions a suboptimum algorithm for evaluating the 2D-CCF potentially increases the radar updating rate while operating with the same computational resources, thus significantly improving the track initiation stage. Alternatively, computational resources saved by the suboptimum algorithm might be exploited to process data coming from parallel channels (e.g., multiple antennas, multiple frequency channels); integrating the obtained results allows recovery of the SNR loss and, potentially, improvement of the target detection performance by exploiting the diversity of information conveyed by the multiple channels, as described in Section 17.6.

The channelization technique gives the option of using only a subset of the original data for computing the final result, such as by discarding some of the frequency channels [27]. This is quite reasonable for many signals of opportunity in which most of the power lies in only a fraction of the sampled bandwidth. This approach would yield additional SNR loss and a degraded range resolution but provides a preliminary estimate of the 2D-CCF by processing a limited amount of data, which might be of great interest when there are tight constraints on the data transfer capability (e.g., in distributed systems). Moreover, channelization based on the DFT is only one such method. Alternative channelization methods might be developed based, for example, on the Walsh transform. Filters based on Walsh transform show higher sidelobes but are computationally cheaper since they require multiplications by ± 1 coefficients.

17.3 | DIRECT SIGNAL AND MULTIPATH/CLUTTER CANCELLATION TECHNIQUES

As is intrinsic in the passive radar concept that involves parasitically exploiting an existing transmitter of opportunity, the characteristics of the transmitted waveform are not under the control of the radar designer and are not tailored for radar application. Because they are used

for transmission of information, the waveforms of opportunity have an intrinsic random behavior. This implies that very often their AFs have time-varying sidelobe structures along both bistatic range and Doppler. Moreover, these sidelobes exist at a level not greatly lower than the peak [3]. This can lead to the following:

1. Strong clutter echoes masking targets with high Doppler frequencies
2. A small fraction of the direct signal being received via the sidelobe/backlobe of the surveillance antenna (still significantly larger than the clutter echo) that masks target echo signals
3. Strong target echoes masking other echoes from other targets of a lower level, even in the presence of large range-Doppler separations

In addition to using carefully designed waveforms, in active radar systems these problems are typically addressed either by applying tapering to the received signal to lower the sidelobes or by using MTI canceller filters to remove the strong stationary clutter echoes. Unfortunately, neither of these can be directly applied to passive radar.

The standard taper functions are typically effective in controlling sidelobes due to the signal characteristics at the leading and trailing edges of the conventional radar pulses that have a well-known shape and can be modified by appropriate filtering. In contrast, the typical sidelobes of the waveforms of opportunity are due to fast amplitude and phase modulations of the waveform spectrum related to the transmitted information content that is obviously unknown and time varying. This makes it impossible to design filters for PBR that are able to control the sidelobes to very low levels.

Also, the masking effect due to the low peak sidelobe level of the AF cannot be removed using conventional MTI techniques. A standard canceller is based on subtracting two or more consecutive echoes (with appropriate weights), with the idea that the echo from stationary targets in two consecutive echoes would be exactly the same and the difference will cancel the undesired contributions from the stationary background. In passive radar, two successive batches of the received signals are different even if they have been reflected by the same stationary background, since the transmitted waveform changes continuously.

Hence, proper adaptive cancellation filters have been designed for this purpose [4, 29–34]. The goal of the cancellation stage is to remove as much as possible the undesired interference to extract the desired target signal from the surveillance signal observation. The complex envelope of the signal received at the surveillance channel can be written as

$$s_{surv}(t) = A_{surv}d(t) + \sum_{m=1}^{N_T} a_m d(t - \tau_{Tm}) e^{j2\pi f_{Dm}t} + \sum_{i=1}^{N_S} c_i d(t - \tau_{ci}) + n_{surv}(t) \quad 0 \leq t < T_0 \quad (17.12)$$

where

T_0 = observation time

$d(t)$ = the complex envelope of the direct signal (a delayed replica of the transmitted signal)

A_{surv} = the complex amplitude of the direct signal received via the sidelobe/backlobe of the surveillance antenna

a_m , τ_{Tm} , and f_{Dm} = the complex amplitude, the delay (with respect to the direct signal), and the Doppler frequency of the m -th target ($m = 1, \dots, N_T$)

c_i and τ_{ci} = the complex amplitude and the delay (with respect to the direct signal) of the i -th stationary ground scatterer ($i = 1, \dots, N_S$); multipath and clutter are modeled as a set of small discrete scatterers; a continuous clutter backscattering environment can be emulated by using a large number of such ground scatterers

$n_{surv}(t)$ = the thermal noise contribution at the surveillance channel

The principle of operation of the adaptive cancellation filter is that, by summing up a number of appropriately delayed and weighted replicas of the transmitted waveform, an estimate of the undesired interference signal is obtained. This estimate is then subtracted from the received signal to remove the interference component.

To follow this approach, the transmitted waveform must be known. The reference signal is thus assumed to be a good replica of the transmitted signal. Specifically, since the direct signal is received by the main lobe of the reference antenna, it is assumed that target and clutter echoes (received from the sidelobes) are negligible. Furthermore, for the sake of simplicity, it is assumed that the reference signal is free of multipath. Under these assumptions, by appropriately filtering the reference signal (a FIR filter with a number of taps emulates the multiple reflections from stationary objects) an adaptive estimate of the undesired contributions in the surveillance signal is obtained. This estimate will then be subtracted from the original signal, leaving an estimate only of the desired target echoes. A widely applied solution for the adaptive cancellation filter is presented in [4], where the joint process estimator algorithm is used. The filter, depicted in Figure 17-6, consists of an adaptive M -stage lattice predictor and an adaptive tapped delay line.

The reference signal samples in Figure 17-6 are first transformed into a corresponding sequence of uncorrelated samples and are then used as the input of the tapped delay line to obtain an estimate of the undesired contributions to be subtracted from the surveillance signal. The gradient adaptive lattice (GAL) is used to adjust the coefficients in the lattice predictor structure, and the normalized least squares (NLMS) algorithm is used to update the coefficients of the tapped delay line.

Both these algorithms are described in details in [35]. The NLMS algorithm is a stochastic gradient algorithm whose free parameters are updated from sample to sample.

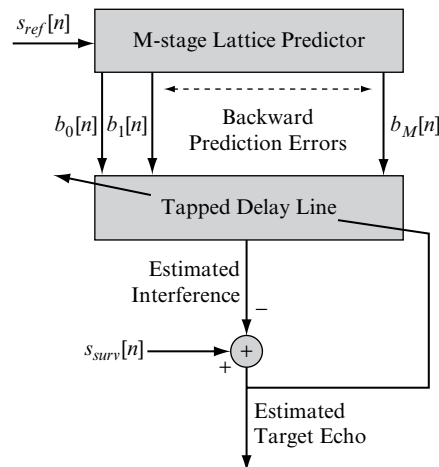


FIGURE 17-6 ■ Sketch of the efficient joint process estimator algorithm.

This algorithm is important because of the simplicity of its implementation (it requires neither measurements of the correlation function nor matrix inversion) and because of its robustness, which is based on the model-independent characteristic of the filter. The main limitation of the NLMS algorithm is its relatively slow rate of convergence. This depends on both the step-size parameter selection and on the eigenvalues of the tap-input data correlation matrix. The GAL algorithm realizes an approximate self-orthogonalization of the input data in the time domain via Cholesky factorization. This can be demonstrated to improve the convergence properties of the subsequent NLMS algorithm, particularly when the eigenvalue spread of the input data correlation matrix is large; this is obtained at the cost of an increase in computational complexity.

The recursive least squares algorithm with sample update and the block least squares algorithm are alternative approaches for disturbance cancellation. These methods realize an exact orthogonalization of the input data in the time domain and hence are characterized by a rapid rate of convergence [35]. A comparative study of such algorithms for PBR purposes is presented in [33].

The block least squares approach has been largely exploited and properly modified to cope with the interference scenarios encountered in different PBR applications [29, 36]. It bears further discussion here.

17.3.1 Extensive Cancellation Algorithm

To derive the basic block least squares algorithm, or the extensive cancellation algorithm (ECA), N samples of the surveillance signal collected during a CPI are arranged in an $N \times 1$ vector:

$$\mathbf{s}_{surv} = [s_{surv}[0], s_{surv}[1], s_{surv}[2], \dots, s_{surv}[N-1]]^T \quad (17.13)$$

Similarly, we collect $N + R - 1$ samples of the reference signal in the following vector:

$$\mathbf{s}_{ref} = [s_{ref}[-R+1], \dots, s_{ref}[0], \dots, s_{ref}[N-1]]^T \quad (17.14)$$

where $R - 1$ is the number of additional reference signal samples to be considered to obtain the desired integration (i.e., N integrated samples) over R time bins.

Assuming that the multipath and clutter echoes are potentially backscattered from the first K range bins, the signal model is exploited by searching for a minimum residual signal power after cancellation of the disturbance (direct signal and ground echoes); thus

$$\min_{\alpha} \left\{ \|\mathbf{s}_{surv} - \mathbf{X}\alpha\|^2 \right\} \quad (17.15)$$

where $\mathbf{X} = \mathbf{B}\mathbf{S}_{ref}$, \mathbf{B} is an incidence matrix that selects only the last N rows of the following matrix, and

$$\mathbf{S}_{ref} = [\mathbf{s}_{ref} \quad \mathbf{D}\mathbf{s}_{ref} \quad \mathbf{D}^2\mathbf{s}_{ref} \cdots \mathbf{D}^{K-1}\mathbf{s}_{ref}] \quad (17.16)$$

whose columns are the zero-Doppler, delayed versions of the reference signal, and \mathbf{D} is a 0/1 permutation matrix that applies a delay of a single sample. The columns of matrix \mathbf{X} define a basis for a K -dimensional disturbance subspace. Solving equation (17.15) yields

$$\alpha = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{s}_{surv} \quad (17.17)$$

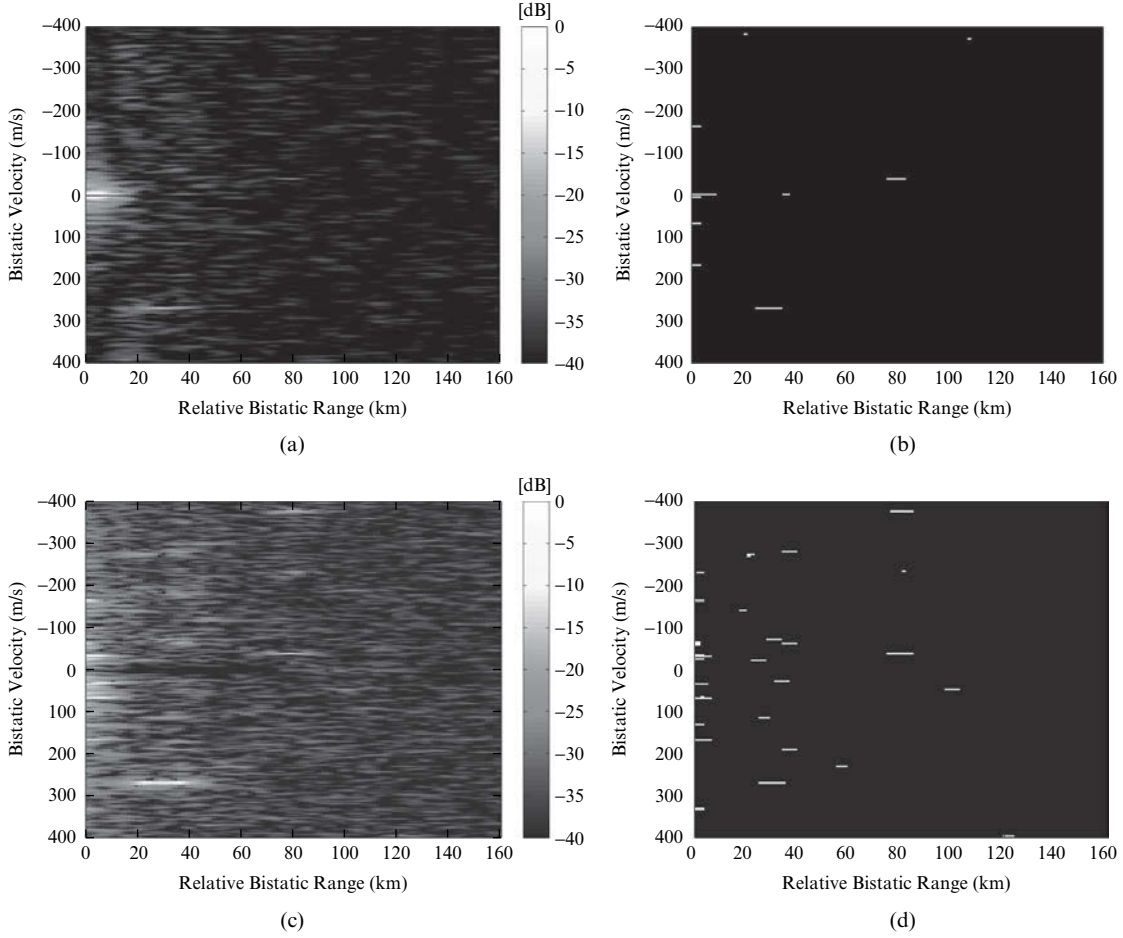


FIGURE 17-7 ■ Detection results for a single data file of 1 sec for a FM-based PBR. (a–b) 2D-CCF and CFAR threshold output ($P_{fa} = 10^{-4}$) after cancellation with ECA for $K = 50$. (c–d) 2D-CCF and CFAR threshold output ($P_{fa} = 10^{-4}$) after cancellation with the batch version of ECA (ECA-B) for $b = 10$ and $K = 50$.

Therefore, the surveillance signal after cancellation becomes

$$\mathbf{s}_{ECA} = \mathbf{s}_{surv} - \mathbf{X}\boldsymbol{\alpha} = \left[\mathbf{I}_N - \mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H \right] \mathbf{s}_{surv} = \mathbf{P} \mathbf{s}_{surv} \quad (17.18)$$

where the projection matrix \mathbf{P} projects the received vector \mathbf{s}_{surv} in the subspace orthogonal to the disturbance subspace.

This algorithm's performance is characterized in [29, 33, 34, 36] against both simulated and real data. In the following some examples are reported against the data recorded with the FM-based PBR prototype described in [29].

Figure 17-7a shows the 2D-CCF over a single recorded data file of 1 second after the application of the ECA with $K = 50$. As expected, a deep and wide null appears at velocity zero, thus allowing a strong target to be easily recognized together with its sidelobe structure.

This is confirmed in Figure 17-7b, where the result of the application of the CFAR threshold to the data is shown using a conventional 2-D cell-averaging CFAR (CA-CFAR)

algorithm with a total number of training bins equal to 120 and a nominal probability of false alarm $P_{fa} = 10^{-4}$ when operating against a zero-mean Gaussian-distributed disturbance. The figure shows that only four of the possible targets are detected. As usually verified in the simulated case, even if significant direct signal cancellation is obtained, target detection might be strongly limited by the disturbance residuals around zero Doppler and the sidelobes of the strongest targets.

In [29] we show that a better disturbance cancellation can be achieved with the ECA by extending the disturbance subspace dimension represented by the columns of matrix \mathbf{X} by including Doppler shifted replicas of the reference signal (i.e., matrix \mathbf{X} with $Q > K$ columns describing a cancellation mask over the range-Doppler plane), thus widening the filter notch in the Doppler dimension. However, this approach is computationally intensive, since it corresponds to increasing the dimension of the weight vector α whose evaluation requires computation and inversion of the matrix $\mathbf{X}^H \mathbf{X}$ with dimensions $Q \times Q$, which corresponds to $O[NQ^2 + Q^2 \log Q]$ complex products.

17.3.2 ECA Batches

A batch version of the ECA (ECA-B) is used to achieve similar or better cancellation performance with limited computational load [29]. Specifically, in this approach, the ECA is applied to a set of b consecutive batches of reduced length $T_b = T_{int}/b$. If the whole temporal extension T_{int} used for coherent integration (2D-CCF) gives rise to $N = T_{int} \cdot f_s$ samples, the ECA algorithm is applied over b consecutive batches of $N_b = N/b$ samples each, as depicted in Figure 17-8. Consider the following for the ECA-B algorithm:

1. At the i -th batch, operating with the same number K of degrees of freedom (DOF), modifying the original ECA will still require evaluating and inverting the matrix $\mathbf{X}_i^H \mathbf{X}_i$

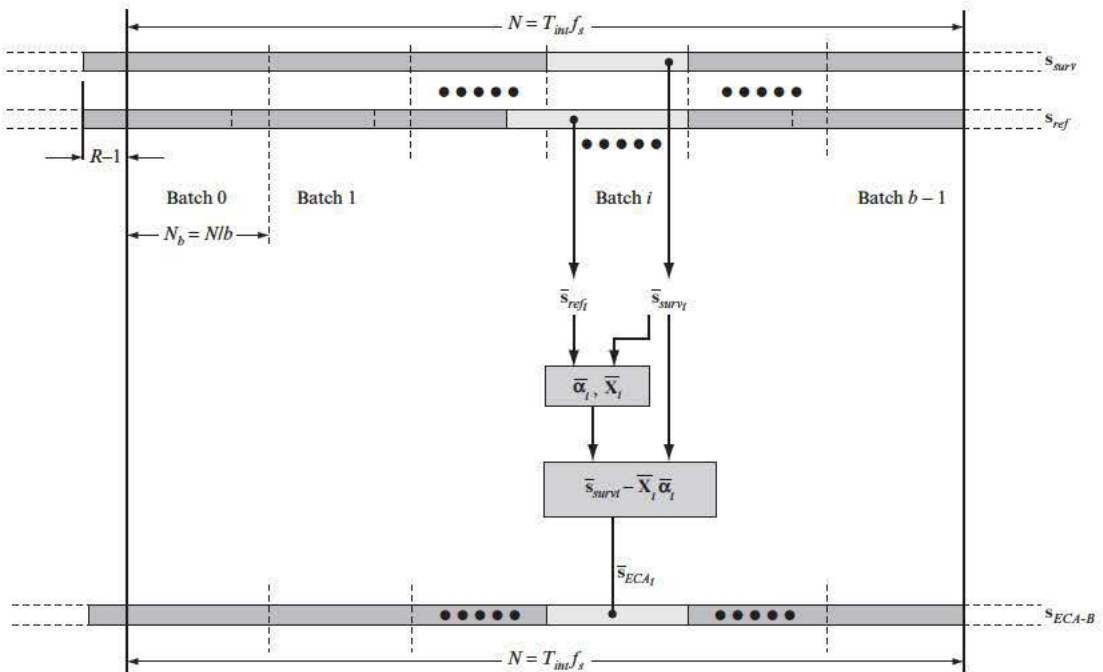


FIGURE 17-8 ■ Sketch of the ECA-B approach.

with dimension $K \times K$, which in this case corresponds to $O[N_b K^2 + K^2 \log K]$ complex products. Clearly this computation needs to be repeated on each batch, thus yielding only a limited increase in the computational burden (about the same order of magnitude in terms of complex products), namely, $O[NK^2 + bK^2 \log K]$ in place of $O[NK^2 + K^2 \log K]$. However, at each batch the dimension of the data can be reduced by a factor b , thus reducing the dynamic storage requirement of the system. In addition, the cancellation processing can run parallel with the data acquisition since it can be started after the first batch has been received. This typically makes the implementation easier and faster.

2. The size N_b of the batch sets the time extent over which the adaptive filter parameters are estimated. Reducing N_b within certain limits increases the adaptivity loss when operating in a stationary environment. However, it allows a certain degree of adaptation to the slowly varying characteristics of typical transmissions, thus making the system more robust.
3. By operating with a reduced temporal extent, T_b , the Doppler resolution capability of the cancellation filter is degraded with respect to the ECA approach operating over the whole T_{int} period. This results in a wider notch in the Doppler dimension that in turn removes the potential request for additional DOF to cope with the spectral dispersion of the interfering signals that would require a higher computational load.

Figures 17-7c–d show the 2D-CCF and the detection results obtained for the same data file of Figures 17-7a–b after the application of the ECA-B for $b = 10$ and $K = 50$. The width of the notch in the Doppler dimension is significantly increased in Figure 17-7c with respect to Figure 17-7a, which was obtained with ECA operating with the same number of DOF. This yields a better cancellation of the disturbance, thus allowing the strongest targets to be easily recognized together with their sidelobe structures. The detection performance has been clearly improved since potential targets are now detected at bistatic ranges greater than 120 km.

Detecting the weakest targets is limited no longer by residuals of the cancellation process but now by the masking effect of the sidelobes of the strongest targets. So it is useless to keep increasing the dimensionality of the disturbance subspace; the DOF saved by using the ECA-B approach can be devoted to counteract this additional limitation. An ad hoc algorithm is presented in [29] and briefly summarized in the following section.

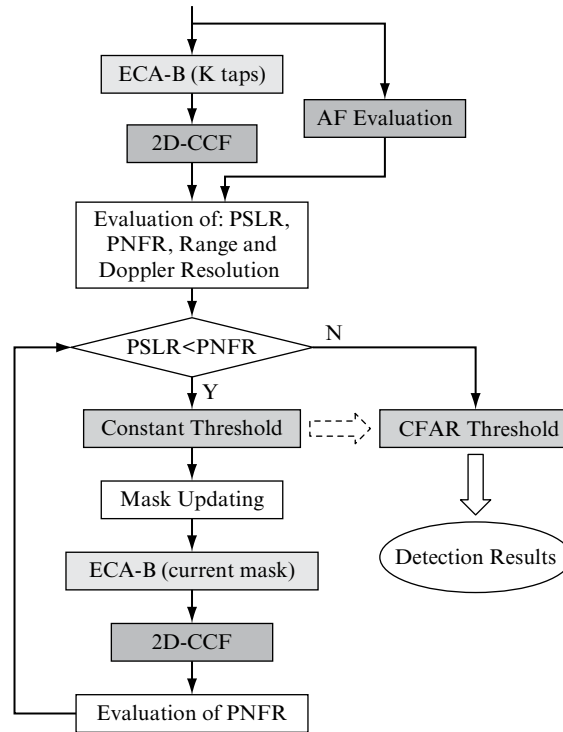
17.3.3 ECA Batches and Stages (ECA-B&S)

A complete processing algorithm is presented, based on the ECA-B approach, that allows:

- a. A preliminary cancellation of the main disturbance contributions
- b. A cancellation performance refinement by the adaptive definition of an extended cancellation mask
- c. The removal of the strongest targets, thus yielding the detection of the weakest ones

The idea of using an iterative cancellation approach is preliminary addressed in [30, 31]. Further, a sequential cancellation algorithm is presented in [34] that exploits a variable number of iterations adaptively identified from the measured data. The approach was designed to prioritize cancellation of the largest disturbance until a desired, predetermined cancellation level was reached. While the ordering strategy and stopping condition

FIGURE 17-9 ■ The ECA-B&S flow diagram.



proposed in [34] allows, in theory, a smart selection of the number of iterations, the application to real data is very time-consuming because of the very high possible number of DOF corresponding to a very high Doppler resolution. However, the complete and very effective algorithm in [29] is based on multiple consecutive stages of the batches and stages version of the ECA (ECA-B&S) and requires only a limited number of iterations to arrive at the optimum results. The main steps are illustrated in Figure 17-9 and are described as follows:

1. The ECA-B is first applied over a range extent including K range bins, and the corresponding 2D-CCF is evaluated. This first step yields near complete cancellation of the disturbance contributions (i.e., direct signal and ground clutter). As previously demonstrated, after this step the disturbance residual power and the strongest targets clearly appear as a set of strong peaks in the 2-D map, together with their sidelobe structures.
2. The AF of the reference signal is evaluated that gives information about (a) range and Doppler resolution, and (b) the peak-to-sidelobe ratio (PSLR). At the same time, using the 2D-CCF the actual noise floor is estimated over a small portion of the Doppler-range surface at high values, where no target is expected and the effects of the waveform (sidelobe structures) may be assumed to be negligible. The peak-to-noise floor ratio (PNFR) is evaluated between the main peak of the 2D-CCF and the estimated noise floor.
3. The PNFR is compared with the PSLR. If the PNFR is smaller than the PSLR, it means that the strongest targets' sidelobes are buried in the noise floor and target detection is limited only by the latter. In this case, a CFAR threshold is applied over the 2D-CCF for a given nominal P_{fa} and the algorithm stops. At this stage, no better target extraction can be expected by further removing detections. To prevent additional stages due to

noise floor estimation inaccuracy, a more conservative stop condition can be used by comparing $\text{PNFR}_{\text{dB}} < \text{PSLR}_{\text{dB}} + \varepsilon_{\text{dB}}$.

4. If the PNFR is higher than the PSLR, a constant threshold is applied on the 2D-CCF from Step 1. This threshold is set at η dB below the main peak of the 2D-CCF, where η is a given value in the range $[0, \text{PSLR}]$ dB. This allows detection of the strongest peaks in the Doppler-range plane due to disturbance residuals and strong targets while avoiding false alarms from sidelobes. Most of the detections associated with disturbance residuals usually appear at very short ranges and can be discarded at a postprocessing stage. The detected target positions are stored and added to the final detection results in a later stage.
5. The Doppler-range bins at the point a detection was declared in Step 4 are used to update the cancellation mask, and a new stage of ECA-B is performed over the extended disturbance subspace basis. This is likely to include the main contributions still limiting the detection of the smaller targets. To obtain an effective cancellation of the strongest peaks, a small area around the corresponding Doppler and range bins is considered for cancellation. The Doppler and range extents can be set adaptively using the resolution values evaluated from Step 2. Subsequently, the resulting 2D-CCF is evaluated and the PNFR value updated.
6. Steps 3–5 are repeated until the algorithm stops. As the number of stages increases, the dimension of the mask over which cancellation is performed grows. As a consequence, the PNFR decreases for two reasons. First, the strongest peaks present in the 2D-CCF at the previous stage have been removed; thus, the actual main peak has a lower absolute value. Second, the noise floor level in the output data increases slightly since a thermal noise contribution is added that comes from the reference signal replicas used for cancellation.

Figure 17-10 shows the results obtained for the same data file of Figure 17-7 at the different stages of the ECA-B&S algorithm (using the parameters listed in Table 17-3).

Three stages of cancellation are performed, within which the cancellation mask is adaptively extended by adding the bins where a peak in the 2D-CCF was found. Specifically, the reference signal AF PSLR is 24 dB, while the estimated PNFR at the first stage is 35.5 dB. Thus, a constant threshold is applied that detects the strong target (Figure 17-10b). At the second stage the estimated PNFR is 32.2 dB (Figure 17-10c), and the constant threshold applied over the 2-D map results in many detections, probably related to disturbance residuals as they are at relative bistatic ranges smaller than 2 km (Figure 17-10d). After the third stage the PNFR decreases to 26.6 dB (Figure 17-10e), and the detection results obtained after the application of the CFAR threshold show an increased number of detections with bistatic ranges of up to 150 km (see Figure 17-10f).

TABLE 17-3 ■ ECA-B&S Parameters
Adopted in Figure 17-10

Parameter	Value
Number of batches b	10
Initial number of DOF K	50
ε	3 dB
η	5 dB
P_{fa} (for the CFAR threshold)	10^{-4}

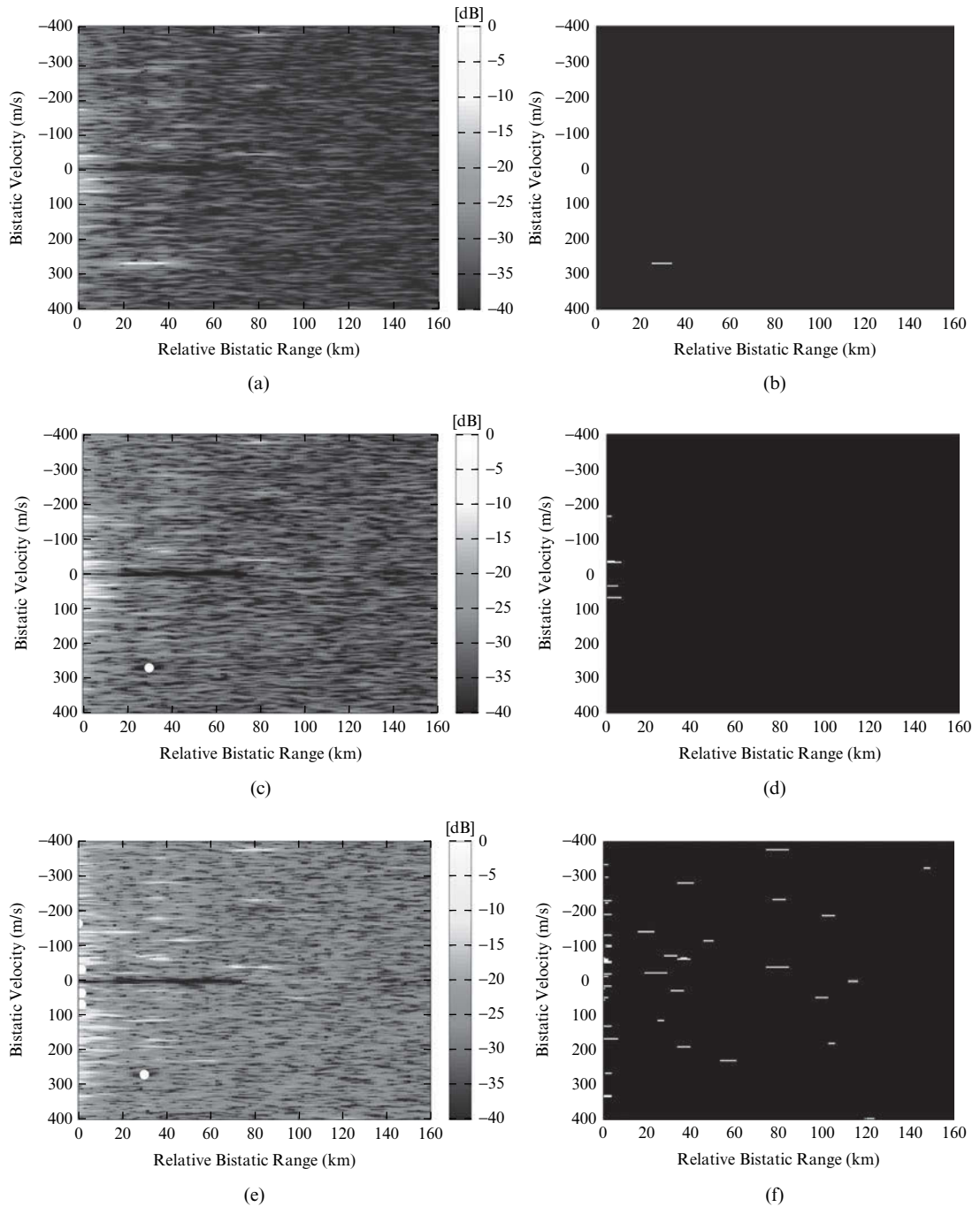


FIGURE 17-10 ■ Results for a single data file of 1 sec for a FM-based PBR with the ECA-B&S algorithm using the parameters listed in Table 17-3. (a)–(c)–(e) show the 2D-CCF after stages 1–2–3. (b)–(d)–(f) show the detection results after stages 1–2–3.

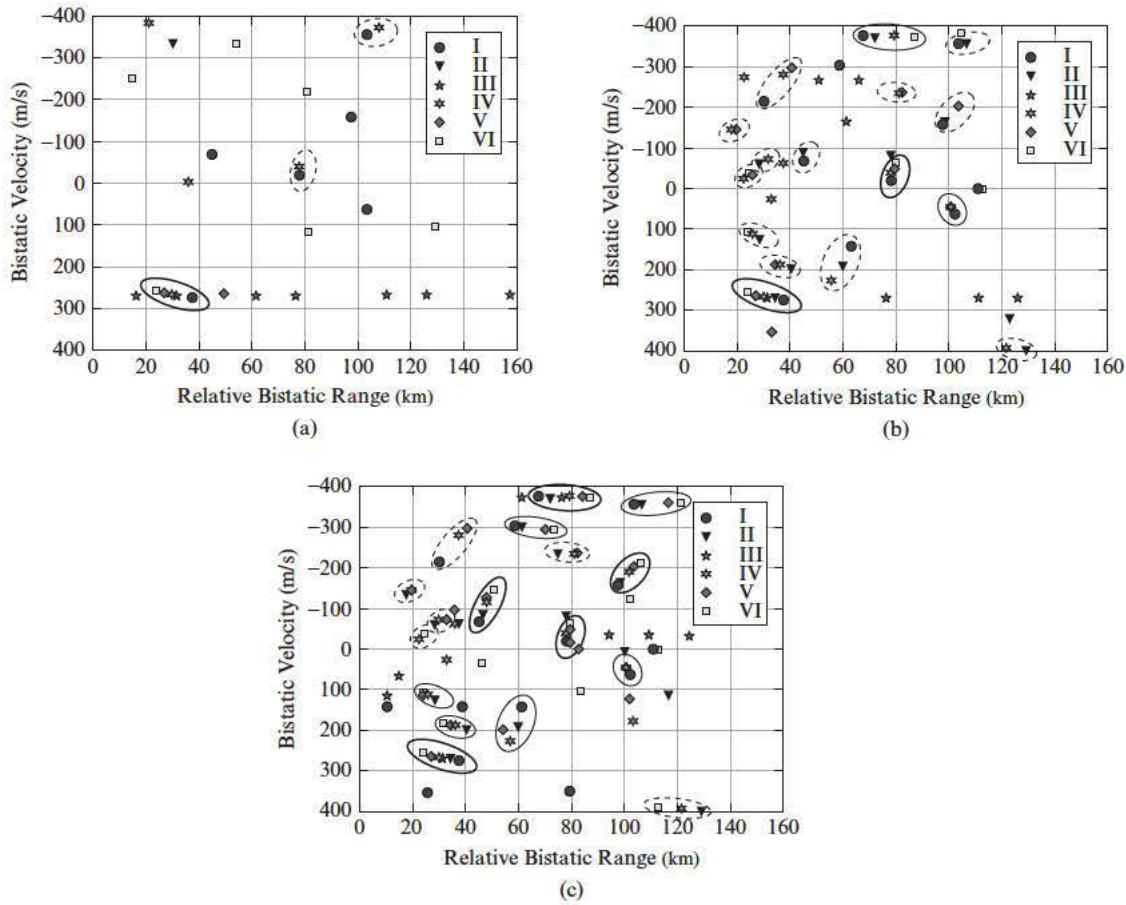


FIGURE 17-11 ■ Superposition of the detection results over the data files sequence obtained with (a) ECA, (b) ECA-B, (c) ECA-B&S.

Figures 17-11a–c show the sequence of the detection results obtained from six consecutive data files using the same range-Doppler map for the three algorithms ECA, ECA-B, and ECA-B&S, respectively. Notice that all the plots within a relative bistatic range of 5 km have been discarded under the assumption that they are due to disturbance residuals. In each figure, the identified plot sequences are indicated with circles using different line styles: the continuous bold line is for sequences containing five to six plots; the continuous thin line is for sequences containing four plots; the discontinuous black is for those containing only three plots; and the discontinuous gray is for possible sequences made of two plots.

When using the ECA (Figure 17-11a), only the strongest target is correctly detected whereas only few plots are generated for other possible targets. Even when the nominal P_{fa} value used for the CFAR threshold is increased, the detection performance does not change but a larger number of false alarms arises (mainly detection of the sidelobes of the used waveforms) [29].

With the ECA-B (Figure 17-11b), detection performance is significantly improved, and many reasonably complete plot sequences are observed; for the others, again only few plots are detected. Also, the third acquisition of the sequence does not yield good detection

performance with any of the algorithms due to the poor characteristics of the particular waveform transmitted at this time. Even in this case, decreasing the detection threshold the target detection capability is not significantly enhanced. However, many false alarms appear, making the target plot sequences more difficult to recognize and hence reducing the capability to track potential targets [29].

When using the ECA-B&S (Figure 17-11c), the completeness of the detected plot sequences is improved since additional plots are detected for a given target track. This gives a significant advantage especially for targets that result in only a two-plot sequence when operating with the ECA-B, since their tracks are otherwise likely to be lost. Moreover, an additional and quite complete sequence appears at about 60 km and -300 m/s that was not identified by the simpler algorithms.

The ECA-B&S yields better detection capabilities with respect to the simpler versions of the ECA and ECA-B. These advantages are obtained at the expense of a higher computational load since additional cancellation stages are required to effectively remove the interference contributions and since the basis of the disturbance space is extended at each stage, including the subspace spanned by the strongest target returns. However, the ECA-B&S allows a reduction in the initial basis dimension of the cancellation algorithm, since the cancellation mask is adaptively updated. Thus, it avoids making the number of DOF of the filter too large. In addition, the faster 2D-CCF evaluation algorithms described in Section 17.2 might be used to speed up the processing at each stage.

Finally, to further reduce computations, a suboptimal implementation of the ECA-B&S is introduced in [29]. The suboptimal implementation of the algorithm performs a cancellation at each stage that involves only the small subspace spanned by the strongest targets detected at the previous stage. This corresponds to an optimistic assumption that this subspace is orthogonal to the subspace spanned by the disturbance basis used in the previous stage. The performed analysis of simulated and real data demonstrates that the two versions of the algorithm are quite comparable in terms of achievable performance, while the suboptimal version allows a dramatic decrease of the computational burden.

17.4 | SIGNAL PROCESSING TECHNIQUES FOR REFERENCE SIGNAL CLEANING AND RECONSTRUCTION

The cancellation and the 2D-CCF evaluation algorithms presented in the previous sections assume the availability of a clean reference signal that corresponds to a high-quality copy of the transmitted signal. Therefore, it might be necessary to perform some transmitter-specific conditioning of the signal that depends on the exploited waveform of opportunity—for example, high-quality analog BP filtering, channel equalization, removal of unwanted structures in digital signals, and complete reconstruction from the received digital signal—before cancellation and 2D-CCF evaluation. Such conditioning approaches are mainly intended for the reference signal aiming at improving its quality (with a benefit for the cancellation step) and the resulting AF (with reduced masking effect for small targets).

Except for the presence of thermal noise in the reference channel, the main degradation to the signal quality is typically related to the presence of multipath. Whereas the

basic processing techniques were introduced with the assumption that no multipath is present, it is easy to imagine that the presence of multipath in the reference signal would yield significant performance degradation. For this reason, different techniques have been proposed to remove such undesired contributions from the reference signal. Obviously, these techniques are waveform dependent and might be based on previous literature for both analog and digital transmissions on channel equalization [37]. Some examples are considered in the following for specific modulation types, among the ones more often exploited for passive radar systems, with the aim of illustrating the problem and offering some useful hints to the design of proper algorithms for reference signal cleaning in PBR.

17.4.1 Constant Modulus Algorithm for Reference Signal Cleaning Using Analog Modulation

To investigate the performance degradation of a PBR exploiting a reference signal affected by multipath, the following model is adopted for complex envelope of the signal collected at the reference antenna:

$$s_{ref}(t) = A_{ref}d(t) + \sum_{m=1}^{N_M} A_m d(t - \tau_m) + n_{ref}(t) \quad 0 \leq t < T_0 \quad (17.19)$$

where we assume that N_M multipath replicas affect the reference signal. In equation (17.19), A_{ref} is the complex amplitude of the direct signal, A_m and τ_m are the complex amplitude and the delay (with respect to the direct signal) of the m -th multipath replica, respectively, and $n_{ref}(t)$ is the thermal noise contribution. Since the direct signal is received by the main lobe of the reference antenna, it is assumed that target and clutter echoes (received from the sidelobes) are negligible.

We now consider a simulated PBR study case where an FM radio broadcast transmission with bandwidth of 150 kHz is exploited. The received signals are sampled at frequency $f_s = 180$ kHz. The signal at the surveillance channel consists of direct signal with a direct signal-to-noise ratio (DNR) of about 61 dB, $N_S = 13$ clutter spikes with clutter-to-noise ratio in the range 40–60 dB and delays of 0.028–0.233 msec, and $N_T = 7$ target echoes with different radar cross sections. The direct signal contribution at the reference channel shows a DNR of about 76 dB; different multipath conditions are now considered.

Figure 17-12a shows the 2D-CCF obtained for the considered study case after applying the ECA operating with $K = 45$ taps in the absence of multipath in the reference signal. The direct signal and the clutter spikes in the surveillance signal have been correctly removed; the obtained clutter attenuation (CA) (i.e., the signal power decrease between the input and the output of the filter) is 54.2 dB. Consequently, all the injected targets are clearly visible as peaks in the 2D-CCF (surrounded by white circles).

When the reference signal is affected by multipath, it is likely that the cancellation capability of the ECA is seriously limited, thus also diminishing its effectiveness for detecting weak targets, whose level is below the clutter sidelobes. This is demonstrated in Figure 17-12b, where a single multipath replica ($N_M = 1$) is present in the reference signal with direct signal-to-multipath ratio (DMR) = 5 dB, delay $\tau_1 = 9/f_s$, and phase shift $\phi_1 = \angle A_{ref} - \angle A_1 = 200^\circ$ between direct signal and multipath echo. A strong peak is present in the 2D-CCF adjacent to the cancellation area, whereas its sidelobes completely mask some targets (discontinuous circles indicate missed detections). The undesired strong peak

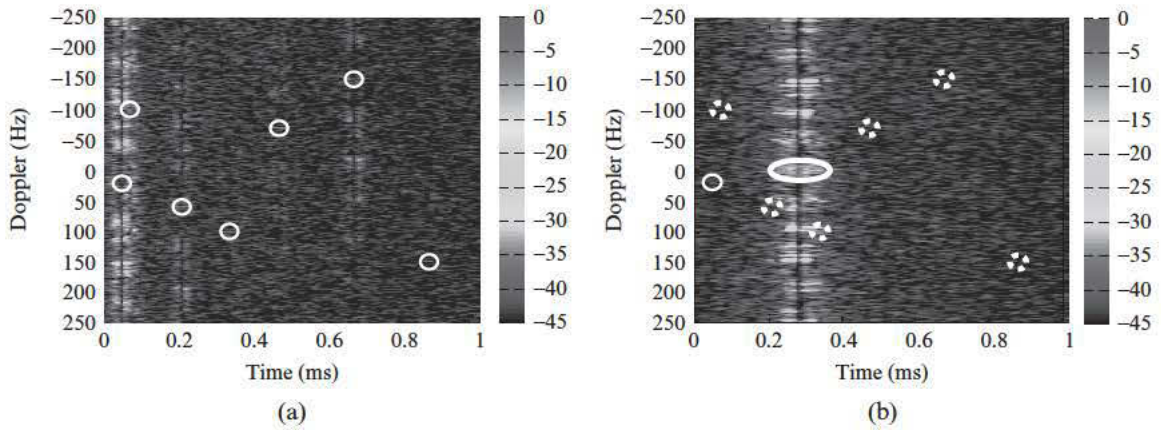
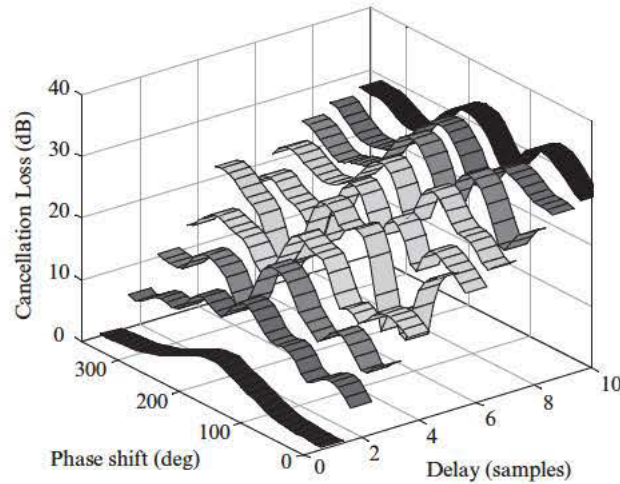


FIGURE 17-12 ■ 2D-CCF (dB) at the output of ECA for a simulated scenario for an FM-based PBR with a reference signal (a) free of multipath, (b) affected by a single multipath replica with $\text{DMR} = 5$ dB, $\tau_1 = 9/f_s$ and $\phi_1 = 200^\circ$.

FIGURE 17-13 ■ CA loss of the ECA for a simulated scenario for an FM-based PBR, with a single multipath replica on the reference signal with $\text{DMR} = 5$ dB.



(indicated by an ellipse) corresponds to the projection of the reference signal multipath echo on the surveillance signal, deriving both from the ECA processing and from the following cross-correlation.

Correspondingly, Figure 17-13 shows the CA loss of the ECA when the reference signal is affected by a single multipath replica with a DMR of 5 dB, as a function of the delay and phase shift of the replica with respect to the direct signal. This CA loss is evaluated with respect to the theoretical CA value, namely, the CA obtained in the absence of multipath in the reference signal (in this study case $\text{CA} = 54.2$ dB). Following the definition of the CA, a high CA loss corresponds to the presence of a strong disturbance residual due to the multipath contribution projection in the surveillance signal whose sidelobes are likely to mask the weak targets present in the scenario. When the multipath echo delay increases, its impact on the CA increases, reducing the cancellation capability of the ECA. In addition, the reduction of the CA depends on the phase shift between the direct signal and the multipath echo; as expected, it is higher when the phase shift is about π .

Thus, proper techniques should be applied to remove the multipath contribution on the reference signal, yielding a pure signal to be used for disturbance cancellation and cross-correlation with the surveillance signal.

Specifically, assuming that the PBR is based on FM radio broadcast transmissions, the well-known constant modulus algorithm (CMA) [38, 39] can be exploited to obtain a blind adaptive equalization of the reference signal. This kind of approach is not limited to our study case, based on FM radio broadcast, but also applies to many of the signals of opportunity available from both analog and digital broadcast transmission systems. Applying CMA to PBR is considered in [40] for an FM-based PBR and in [13] for a GSM-based PBR, where only a basic temporal version of the algorithm is applied. References [41, 42] consider the possibility of using a small antenna array for the reception of the reference signal from the transmitter of opportunity, and two additional versions of the algorithm are used for PBR application along the line suggested in [43–46]: the space-CMA (S-CMA) that operates on the signal samples collected at the different elements of the array; and the space-time-CMA (ST-CMA) that operates on the delayed samples of the signals collected at the multiple channels.

The CMA tries to suppress the additive interference at the input signal by constraining its output to be a constant modulus signal, accomplished by minimizing the cost function

$$J = \frac{1}{4} E \{ [|y[n]|^2 - 1]^2 \} \quad (17.20)$$

where $E\{\cdot\}$ is the statistical expectation, and $y[n]$ is the CMA output at the n -th time instant. For the ST-CMA it is given by

$$y[n] = \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} w_{m,l}[n] s_{ref,m}[n-l] \quad (17.21)$$

where $w_{m,l}[n]$ is the l -th adaptive weight to be applied on the m -th channel at the n -th time instant, and $s_{ref,m}[n]$ is the reference signal at the m -th channel at the n -th time instant ($m = 0, \dots, M-1; l = 0, \dots, L-1$). The corresponding processing scheme is depicted in Figure 17-14. The filter weights to be applied are adaptively updated according to the following strategy:

$$w_{m,l}[n+1] = w_{m,l}[n] - \mu \varepsilon[n] s_{ref,m}^*[n-l] \quad (17.22)$$

where

$$\varepsilon[n] = \{|y[n]|^2 - 1\} \cdot y[n] \quad (17.23)$$

and μ is the step-size parameter that determines the rate of convergence and the accuracy of the algorithm. Obviously, using a single-antenna element ($M = 1$), we come back to the original time-only CMA (T-CMA) that operates as a FIR filter with L taps in the temporal dimension. The corresponding filter tries to invert the estimated transfer function of the propagation channel. Using a single tap for each antenna element ($L = 1$) a space-only version of the algorithm is obtained (S-CMA) that tries to estimate the direction of arrival (DOA) of the multipath replicas because of the spatial discrimination capability given by the receiving array. The filter modifies the quiescent antenna pattern by synthesizing an equivalent pattern in which zeros are set at given angles corresponding to the DOA of the multipath echoes. Finally, considering multiple antenna elements and

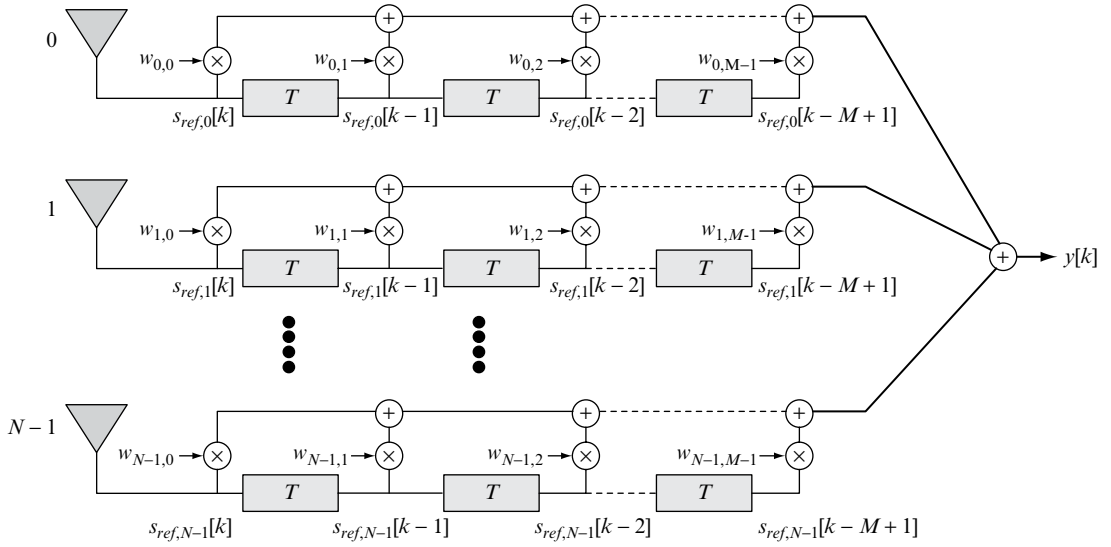


FIGURE 17-14 ■ Sketch of the space-time CMA approach.

filter taps, a fully ST-CMA is obtained that is expected to operate as a combination of the previous two versions (T-CMA and S-CMA) by contemporaneously using the temporal and spatial DOF.

An extensive performance comparison among the different monodimensional CMA versions shows that both approaches allow effective equalization of the reference signal in a number of different multipath scenarios [41]. However, both approaches also have been shown to yield significant loss in particular situations. Specifically, the T-CMA yields good performance for a limited multipath replica's delay, while the S-CMA yields an effective equalization only when the multipath echo DOA is far enough from the direct signal's DOA. Besides these considerations, the performance of the different versions of the CMA for multiple multipath replicas affecting the reference signal strongly depends on the particular phase shift combination between the replicas. The performance of the monodimensional CMA versions might be slightly improved increasing the number of DOF in the temporal and spatial dimension, respectively; however, this is paid in terms of computational load and system complexity. Nevertheless, a residual loss still might be experienced in certain scenarios even if significantly increasing the number of DOF. This is especially evident when a certain degree of a priori knowledge on the multipath scenario is not available or is time varying. It is then apparent that the joint exploitation of the temporal and spatial DOF can be of great potential interest for multipath removal on the reference signal exploited by a passive radar. The analysis demonstrated that in most cases the ST-CMA correctly equalizes the reference signal, thus yielding a pure reference signal to be used for the following cancellation stage; moreover, thanks to the advantage of a space-time processing, the ST-CMA has been shown to be robust even when operating with a small number of antennas.

As an example, Table 17-4 reports the cancellation loss obtained after the application of all the considered CMA versions in different multipath scenarios (T-CMA with $L = 100$, S-CMA with $M = 4$, and ST-CMA with $L = 100$ and $M = 4$). Also the CA loss obtained without any multipath cancellation algorithm is shown for comparison. Specifically, two

TABLE 17-4 ■ CA Loss (dB) in Different Simulated Scenarios for an FM-Based PBR, for Two Multipath Replicas Affecting the Reference Signal with DMR = 5 dB

Study Case		No CMA	T-CMA	S-CMA	ST-CMA
Both replicas with limited delays and high DOA values	$\tau_1 = 1/f_s$ $\theta_1 = 10^\circ$	6.9	0.1	0.0	0.0
	$\tau_2 = 5/f_s$ $\theta_2 = -15^\circ$				
Replicas with high delays and high DOA values	$\tau_1 = 10/f_s$ $\theta_1 = 20^\circ$	36.2	14.9	0.0	0.0
	$\tau_2 = 9/f_s$ $\theta_2 = -10^\circ$				
Replicas with high delays and low DOA values	$\tau_1 = 10/f_s$ $\theta_1 = 2^\circ$	33.0	19.9	45.4	0.0
	$\tau_2 = 2/f_s$ $\theta_2 = -4^\circ$	35.3	54.0	27.6	0.0
	$\tau_1 = 6/f_s$ $\theta_1 = 2^\circ$ $\tau_2 = 10/f_s$ $\theta_2 = -3^\circ$				

multipath replicas (DMR = 5 dB), with different combinations of their delays τ_m and DOA θ_m spanning four possible study cases (both replicas show the same phase shift $\phi_1 = \phi_2 = 0^\circ$), are analyzed. When the replicas show limited delays and high DOA values, all the CMA versions are able to recover the cancellation loss obtained without prefiltering the reference signal. For multipath replicas with higher delays, only the CMA versions that exploit spatial DOF allow the reference signal to be effectively equalized. However, when the replica delays are high and their DOA are close to the direct signal's, the monodimensional versions of the CMA become unfeasible, yet the ST-CMA is still able to yield a quite clean reference signal and negligible cancellation loss on the surveillance channel.

Figures 17-15a-c show the 2D-CCFs obtained after the application of the different CMA versions when the same study case of Figure 17-12 is considered for the surveillance signal and the fourth multipath scenario of Table 17-4 for the reference signal. In the unfavorable situation previously described, both the T-CMA and the S-CMA do not allow any of the injected target to be detected. In contrast, by contemporaneously using the temporal and spatial DOF, the reference signal is effectively reconstructed, allowing the disturbance contribution in the surveillance channel to be properly removed and, consequently, the injected targets to be detected. The result shown in Figure 17-15c is very similar to that obtained in the absence of multipath on the reference signal (see Figure 17-12a).

17.4.2 Reference Signal Reconstruction for Digital Transmissions

When digital transmissions are exploited by a PBR, the multipath removal from the reference signal can be performed by demodulating the signal received at the reference antenna. A quite good copy of the transmitted signal can then be obtained by remodulating the resulting symbol sequence according to the transmission standard [6].

A few considerations are in order about the required demodulation process. In principle, the symbol sequence extraction can be arrested at any desired layer within the network protocol. Obviously, the PBR performance is mainly affected by the physical layer characteristics; thus, extracting the symbol sequence should involve at least the demodulation of the received signal and the channel equalization. This can be obtained via well-known algorithms developed for digital communication [37], whose complexity depends on the considered transmission standard, the required accuracy in the extraction process, and the severity of the considered scenario. Better signal reconstruction might be obtained by exploiting the upper layers of the network. This would require the symbol sequence to

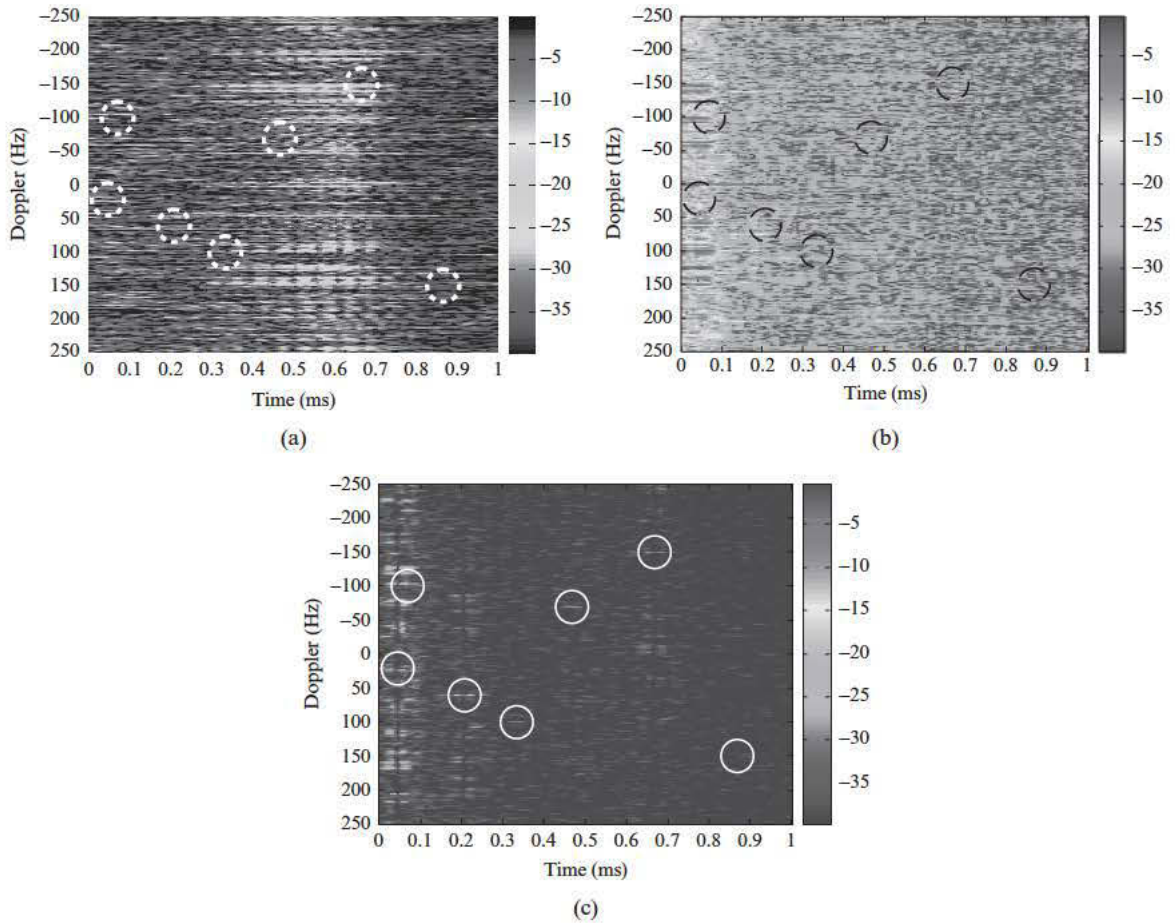


FIGURE 17-15 ■ 2D-CCF after applying the ECA on the surveillance signal and different CMA approaches on the reference signal for the considered study case and the fourth multipath scenario of Table 17-4. (a) Time-CMA. (b) Space-CMA. (c) Space-time-CMA.

be decoded up to the extraction of the original sequence of information bits and allows us to exploit the redundancy introduced intentionally at each layer within the network protocol to correct potential errors. This advantage is obviously paid in terms of higher computational complexity; moreover, it requires perfect knowledge of the standard for each layer. Typically, extracting the symbols sequence at the physical or medium access control layers provides a quite good reconstruction of the transmitted signal. Also, if we are not interested in the real transmitted data sequence, differential demodulation may be adopted for phase-shift keying (PSK) transmissions because this allows to remove difficulties encountered when implementing coherent demodulation [6].

As previously mentioned, the PBR principle is based on the availability of a good copy of the transmitted signal to be used for both disturbance cancellation and 2-D matched filtering. Recovering the correct transmitted symbol sequence and repeating the ideal modulation process do not necessarily lead to a perfect reconstruction of the transmitted signal. In fact, the ideal reconstruction of the transmitted signal by remodulating the extracted symbol sequence does not account for the signal modifications induced by

the transmitter (e.g., band-pass filtering adopted to match the standard specifications, carrier frequency drift and oscillator phase noise, additional undesired modulations). Such effects are usually removed in the demodulation process, together with the transmission channel effects, since they introduce degradation in the extraction of the transmitted symbol sequence. In contrast to the case of the communication problem, where the information content is of interest, the objective of the reconstruction process for a PBR is to obtain an exact copy of the transmitted signal that includes the above mentioned transmitter-induced signal modifications. If these features are lost, their absence causes a decorrelation between the reconstructed signal and the signal received at the surveillance channel. This must be partially solved by postprocessing the ideal reconstructed signal with the aim to apply the same modifications introduced by the transmitter (in primis the shape of the passband filters), which can be either a priori known or estimated from the received signal.

In digital transmissions, an alternative and highly attractive approach is possible. In fact, the reference signal may be synthesized directly from the surveillance signal without requiring a dedicated receiving channel. This can be obtained by demodulating the signal collected at the surveillance channel to extract the sequence of transmitted symbols, when it is possible to assume that the direct signal from the transmitter is by far the strongest contribution received in the surveillance channel. As illustrated in the introduction, this is often the case since the direct signal (despite the fact that it is received through the sidelobes) is typically much higher than the low-level signal reflected from any moving target. In these conditions, the reconstruction of the symbols sequence potentially yields only a few errors even for SNR values that are not very high. In addition, as in the previous case, proper equalization techniques and optimized demodulation and decoding schemes might be exploited for error correction and better recovery of the transmitted signal. This approach in principle eliminates the effects of other signal contributions such as multipath and target echoes. Thus, a pure reference signal may be obtained by remodulating the extracted symbol sequence according to the standard specifications and inserting the additional transmitter-specific features, as discussed already.

As an example, we report the experimental results obtained for a WiFi-based PBR using the experimental setup described in [18, 21]. For this target detection experiment, the transmitter of opportunity, a wireless access point (AP), was set up to roam for connected devices emitting the beacon signal at 1 ms intervals. A quasi-monostatic configuration was adopted for the two antennas mounted one on top of the other. Two targets were present (a car and a running man) moving away from the antenna location at approximately 25 km/h and 20 km/h, respectively.

Specifically Figures 17-16a–b show the 2D-CCF evaluated, before and after the disturbance cancellation stage, with an integration time of 0.4 sec. These figures were obtained with a pure reference signal directly collected at the output of the AP employing a directional coupler to gather a fraction of the transmitted signal in a dedicated receiving channel (reference channel). The 2D-CCF has been normalized to the nominal thermal noise power level; the value at each map location represents the estimated SNR. The reported map has been upper-saturated to a value of 50 dB so that the same color scale could be used for all the reported results for a fair direct comparison (even after disturbance cancellation). A strong peak appears in Figure 17-16a at zero range and zero velocity that corresponds to the direct signal transmitted by the transmitting antenna and collected by the sidelobes of the surveillance antenna. Furthermore, all the 2D-CCF cut at zero velocity is characterized by very high SNR values (well above the limiting color scale value of 50 dB) due to

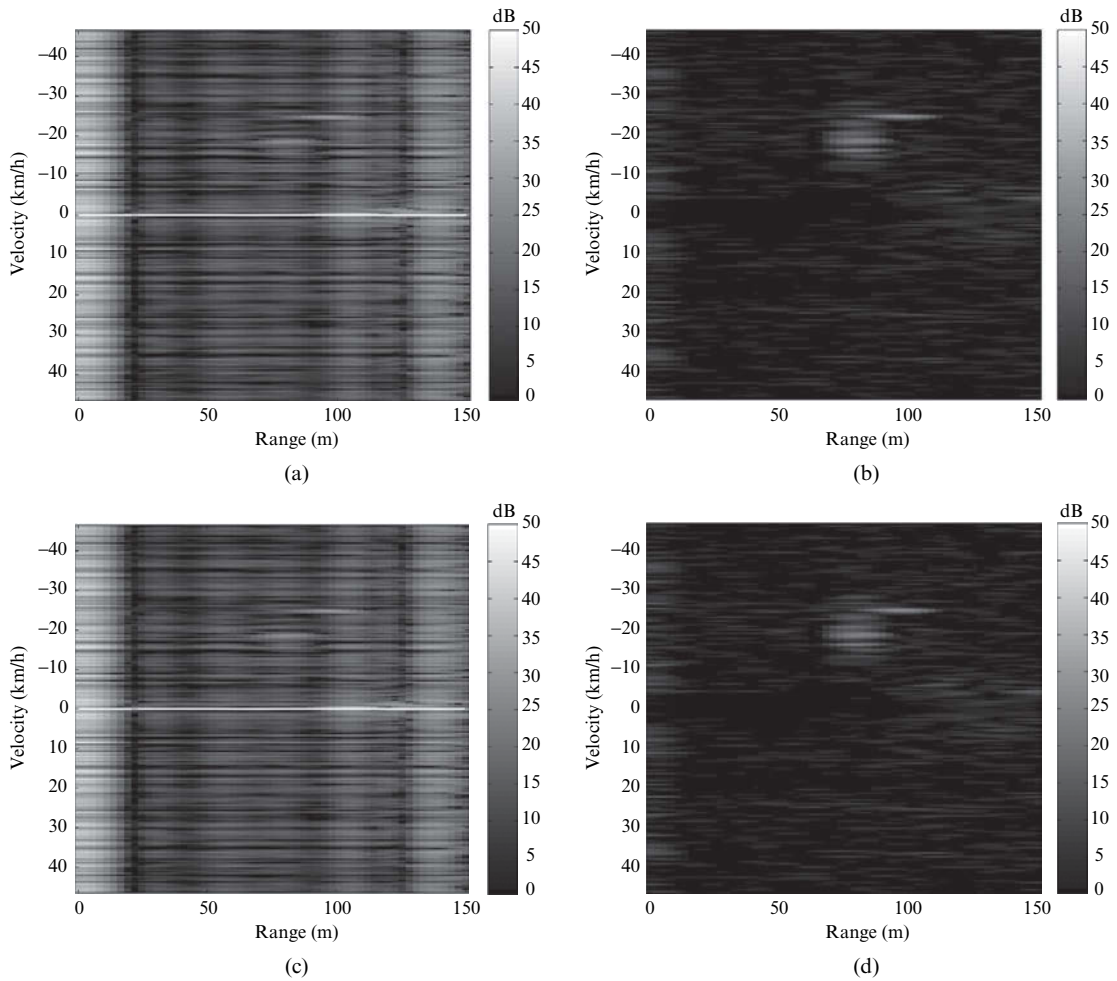


FIGURE 17-16 ■ 2D-CCF for a WiFi-based PBR. (a) Without disturbance cancellation using the registered version of the transmitted signal. (b) With disturbance cancellation using the registered version of the transmitted signal. (c) Without disturbance cancellation using the recovered version of the transmitted signal. (d) With disturbance cancellation using the recovered version of the transmitted signal.

both the range sidelobes related to the direct signal contribution and the reflections of the transmitted signal from stationary scatterers. Each peak at zero velocity comes with the corresponding Doppler sidelobe structure, and this yields a quite high sidelobe level all over the 2D-CCF that completely masks the two targets returns. Figure 17-16b reports the 2D-CCF evaluated after disturbance cancellation but without sidelobes control. The cancellation filter (ECA-B) operates over consecutive batches of duration $T_b = 30$ ms. The filter length was set to remove reference signal replicas with delays corresponding to ranges between 0 and 330 m. The direct signal and all disturbance contributions at zero Doppler are effectively removed, thus significantly increasing the useful dynamic range for target detection. The car (moving at -25 km/h) now appears as a strong peak in the map, and the human target return is significantly spread in the Doppler dimension around its actual location (range = 85 m, velocity = -20 km/h). This is in a measure due the

superposition of the scattering contributions from the different parts of the body moving in different manners.

The same analysis is then repeated by using only the signal received in the surveillance channel, from which the direct signal is also reconstructed. The corresponding results are reported in Figures 17-16c–d using the recovered version of the transmitted signal obtained by demodulating and remodulating the surveillance signal. The results obtained with this approach are largely comparable to those obtained with the reference signal directly collected at the output of the AP. The different tests performed showed that the detection capability of the considered system remains almost unchanged; the few differences are from the impossibility of reintroducing the exact behavior of the transmitter and the receiver, thus including their selectivity and frequency accuracy.

The proposed approach has a great interest in a wider range of PBR systems using digital waveforms (e.g., DVB-T, DAB, GSM signals). In fact, when the reference signal can be directly recovered from the signal collected at the surveillance antenna, a dedicated receiving channel can be eliminated, thus significantly reducing the system complexity. Nevertheless, when the transmitted signal has to be collected by a directive antenna or a dedicated beam synthesized from an electronically steerable array of antennas to guarantee an acceptable SNR, the proposed approach is still able to significantly improve performance since it might yield a cleaner version of the reference signal with reduced multipath or other disturbance contributions. In this regard, this approach might be a successful strategy for recovering the transmitted signal in single-frequency networks like those for DVB-T and DAB broadcast transmissions.

17.5 | 2D-CCF SIDELOBE CONTROL

PBR operation inherently implies that the transmitted waveform is not within the control of the radar designer. This contrasts with the usual case of conventional radar systems, where the transmitted waveform is carefully designed to provide an AF with appropriate properties (e.g., narrow peak in both range and Doppler and low sidelobes). The implications include bad behavior for the sidelobes of the 2D-CCF with specific features slightly different in the cases of analog and digital transmissions.

When analog transmissions are exploited (e.g., in an FM-based PBR) due to the variable and unpredictable characteristics of the transmitted FM waveform, the sidelobes of the AF usually have a time-varying structure and exist at a level not greatly lower than that of the peak [3, 47]. Thus, target returns are likely to be masked by the high level sidelobes resulting from the disturbance contributions (direct signal breakthrough and multipath reflections), even in the presence of a significant range-Doppler separation and proper techniques should be applied to remove such undesired contributions, as previously described.

When digital transmissions are used, the situation becomes even more severe. In fact, the presence of specific periodic or pseudo-periodic features in the modulation format of the signal of opportunity leads to a number of undesired peaks in the 2D-CCF. Possible examples of such specific features are the chipping sequence in a direct sequence spread spectrum (DSSS) modulation [16, 18], the repetition of the symbols in the guard intervals and the presence of pilots subcarriers in an OFDM modulation [6-9], and the multiple repetitions of preassigned symbol sequences usually exploited to obtain synchronization in many digital standards. The corresponding sidelobes structures present in the 2D-CCF

of these signals might be responsible for severe masking effect on weak targets or for the increase of the false alarm rate, thus significantly reducing the target detection capability of the PBR system.

Fortunately, waveforms with digital transmissions have a deterministic behavior and their sidelobes can be effectively reduced if a priori knowledge of their position and intensity is available.

17.5.1 Range Sidelobes Control for 2D-CCF in WiFi-Based PBR

In local area monitoring applications, which aim to detect and localize designated human beings or man-made objects within short ranges, wider bandwidth signals of opportunity should be used to achieve the required range resolution. Thus, wireless local area network transmissions might be considered. Wireless networking applications are proliferating at a very rapid rate for both commercial and private use. One of the most popular systems is the IEEE 802.11 standards-based technology (commonly known as WiFi), which offers reasonable bandwidth (range resolution) and coverage and wide accessibility.

WiFi transmissions are complex. A number of standards have been developed, and further versions are being investigated [48], the three most commonly of which are 802.11a, 802.11b, and 802.11g. 802.11a is adopted only in the regulation domain of the United States and operates in the 5 GHz band, whereas 802.11g represents the third generation of wireless networking standards (after 802.11 and 802.11b) and is fully compatible with the older standards operating in the 2.4 GHz band.

A theoretical analysis of the WiFi signal AF is in [16] with reference to different transmission standards. Also, some experimental results have been reported in the literature showing the potentialities of a WiFi-based PBR, [18–21]. The processing algorithms described in the previous sections might be easily applied to a WiFi-based PBR. The transmission is of pulsed type, so 2D-CCF evaluation and adaptive weight update in the cancellation filter might be synchronized to the AP duty cycle. In this regard, the batches algorithm for evaluating 2D-CCF is a natural choice and goes together with setting the duration of the batch equal to the pulse length (and neglecting the Doppler compensation within each transmitted pulse):

$$\chi[l, k] \cong \sum_{m=0}^{N_P-1} e^{-j2\pi \frac{k i_m}{N}} \chi^{(m)}[l] \quad (17.24)$$

where $\chi^{(m)}[l]$ is the l -th sample of the temporal cross-correlation function (CCF) among the surveillance and the reference signal evaluated over the m -th pulse, N_P is the number of pulses within the CPI, and i_m is the time index corresponding to the first sample of the m -th pulse.

According to the physical layer specifications adopted by each standard, in typical operation modes a WiFi AP may be broadcasting a mix of signals based on different modulations and coding schemes that strongly affect the corresponding autocorrelation function (ACF). The main modulations are either DSSS or OFDM, and data rates between 1 and 54 Mbps are currently specified. DSSS is the most common modulation, with OFDM dominating only at higher data rates. In addition, an IEEE 802.11 AP periodically transmits a regular beacon signal broadcasting its presence together with channel information. To allow stations capable of different modulation schemes in a wireless LAN to coexist, the AP usually uses the DSSS modulation for the beacon frames.

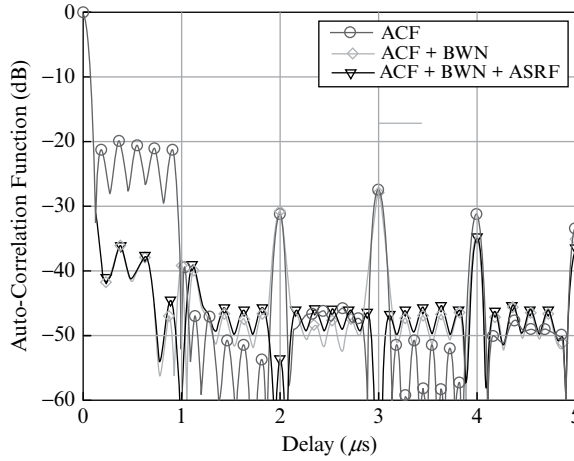


FIGURE 17-17 ■
ACF with sidelobes
control filters for a
DSSS WiFi signal
registration of
duration 0.25 s.

The DSSS system provides a processing gain of at least 10 dB that is accomplished by chipping the baseband signal at 11 MHz with an 11-chip pseudo noise code. An 11-chip Barker sequence is used as the pseudo noise code sequence, $\mathbf{c} = [1, -1, 1, 1, -1, 1, 1, 1, -1, -1, -1]$, where the leftmost chip is output first in time. The first chip is aligned at the start of a transmitted symbol so that each symbol duration is exactly 11 chips long ($T_{\text{sym}} = 11 \cdot T_{\text{chip}} = 1 \mu\text{s}$). Consequently, the resulting signal ACF will be dominated by the Barker code [16].

The ACF obtained for a DSSS WiFi signal registration of duration 0.25 s is reported in Figure 17-17, for the AP broadcasting beacon frames at reasonably regular intervals. Aiming at a local area surveillance application, a delay extent of $5 \mu\text{s}$ is considered that corresponds to a bistatic range extent of 1.5 km. Specifically, five almost identical sidelobes appear within time delays of $T_{\text{sym}} = 1 \mu\text{s}$ (300 m) that are obviously due to the exploited 11-chip Barker code. As expected, these sidelobes show a PSLR of about $20\log_{10}(11) = 20.83 \text{ dB}$ and a temporal separation equal to $2T_{\text{chip}}$. In addition, other strong sidelobes appear at delays multiple of $T_{\text{sym}} (k \cdot 300 \text{ m})$. These sidelobe structures are due to the cyclical repetition of the Barker code, and their level equals the average correlation among consecutive symbols of the data frame. Thus, assuming a random transmission and reasonably large integration times, the corresponding PSLR may be significantly greater than that due to the Barker code sidelobes. However, such additional sidelobes structures should not be a priori neglected since their level is highly time varying and strongly depends on the signal content. For example, the PSLR could be seriously degraded if identical pulses of limited duration are periodically transmitted, as in the case of the beacon frame. In fact, even increasing the integration time, the actual cross-correlation among consecutive symbols is lower bounded by the cross-correlation evaluated over the single pulse consisting of a limited number of symbols.

Therefore, proper filters are presented in the following for the sidelobes level control distinguishing between those deriving from the Barker code and those appearing at multiples of T_{sym} .

Weighting Network for Barker Sidelobes Reduction (BWN)

An effective weighting network for reducing the Barker sidelobes in the DSSS WiFi signal ACF is presented in [18] along the lines proposed in [49]. Its design exploits the a priori

TABLE 17-5 ■ PSLR for Different BWN Lengths

Filter Length (L_{BWN})	PSLR (dB)
25	26.5
51	34
101	50
151	63

knowledge of the 11-chip Barker code ACF:

$$\alpha_B[n] = \sum_{k=0}^{10} c[k]c[k+n] = \begin{cases} 0 & n \text{ odd} \\ 11 & n = 0 \\ -1 & n \text{ even} \end{cases} \quad (17.25)$$

The BWN is implemented as a transversal lattice filter with impulse response $h_{BWN}[n]$ and total filter length L_{BWN} . Indicating with $\alpha_{BWN}[n] = \alpha_B[n] * h_{BWN}[n]$ the output of this weighting network applied to the 11-chip Barker code ACF, the filter coefficients might be found by solving the following constrained optimization problem:

$$\begin{cases} \alpha_{BWN}[0] = \max \\ |\alpha_{BWN}[n]| \leq 1 \end{cases} \quad (17.26)$$

To find the solution, each modulus inequality of the system in (17.26) is replaced by two linear inequalities, and well-known linear programming algorithms are used. To this purpose we exploited the CVX MATLAB-based modeling system for convex optimization [50]. By properly selecting the filter length L_{BWN} , different values of the PSLR can be achieved as reported in Table 17-5.

The BWN can be applied to the CCF among the surveillance and the reference signal evaluated on a pulse basis before coherent integration over a train of consecutive pulses (see equation (17.24)). To obtain an alternative implementation, the cross-correlation can be evaluated between the surveillance signal and a BWN prefiltered version of the reference signal. It is to be noticed that, if the filter length is assigned by the specific surveillance application requirements, the optimization required for filter weights evaluation has to be performed only once.

The theoretical performance of the BWN has been analyzed in [18] against simulated data. Here, the result of the application of the BWN with total length $L_{BWN} = 101$ is reported in Figure 17-17, compared with the original ACF for a real signal registration of duration 0.25 s. As expected, the BWN dramatically decreases the ACF sidelobes due to the Barker code. However, the BWN is not able to obtain the theoretical PSLR of 50 dB; this can be explained by observing that the original ACF shows some deviations with respect to its theoretical version due to the AP and the receiving channel filters that reduce the actual signal bandwidth with respect to its nominal value. Nevertheless, notice that a PSLR greater than 35 dB is obtained that results in a significant increase of the useful dynamic range.

Additional Sidelobes Reduction Filter (ASRF)

While the conceived BWN largely reduces the Barker sidelobes, it is ineffective against the periodic sidelobe structure due to the nonzero correlation among consecutive symbols of the beacon frame. Aiming at the detection of targets at delays higher than $0.8 \mu\text{s}$ (ranges

higher than 240 m), the temporal ACF sidelobes at multiples of $1 \mu\text{s}$ should be reduced to avoid masking effects on targets appearing at those delays. A proper strategy to counteract this effect is presented in the following.

First, whereas the positions of these additional peaks in the ACF are a priori known, their value is data dependent and has to be estimated from the data themselves. To this purpose the reference signal can be used to evaluate the values of the actual signal ACF at delays $t_k = kT_{\text{sym}}(\hat{\alpha}(t_k))$; these estimates can be then exploited to design a simple and effective sidelobe reduction filter. Specifically, we start from the original CCF between the surveillance and the reference signals for the m -th pulse, $\chi^{(m)}(\tau)$. The required sidelobes reduction can be obtained by subtracting from $\chi^{(m)}(\tau)$ a number $2K_S$ of its replicas, appropriately scaled and delayed:

$$\chi_{\text{ASRF}}^{(m)}(\tau) = \chi^{(m)}(\tau) - \sum_{\substack{k=-K_S \\ k \neq 0}}^{K_S} \gamma_k^{(m)} \chi^{(m)}(\tau - kT_{\text{sym}}) \quad (17.27)$$

where K_S is the number of range sidelobes to be removed on both sides of the ACF main peak, and $\gamma_k^{(m)}$ is the scaling factor for the k -th sidelobe at the m -th pulse that is obtained as

$$\gamma_k^{(m)} = \frac{\hat{\alpha}^{(m)}(kT_s)}{\hat{\alpha}^{(m)}(0)} \quad (17.28)$$

The number K_S of sidelobes to be reduced can be arbitrarily selected aiming at removing periodical sidelobes contribution over a given range extent corresponding to the surveillance area. The ASRF can be applied after the BWN by properly modifying the scaling factors, namely, the samples of the BWN-filtered reference signal ACF should be used.

The ACF resulting from the cascade of the two different sidelobe reduction filters is shown in Figure 17-17 when the ASRF operates with $K_S = 3$. The joint application of the two considered approaches in the range dimension provides a significant improvement of the PSLR and the effective removal of the isolated peaks, with only a very limited loss in SNR that is mainly due to the BWN.

To show the benefits of the proposed techniques with reference to the target detection issue, Figure 17-18 reports the results obtained for the same data set of Figures 17-16a–b after applying the range sidelobes control networks (BWN + ASRF).

Figure 17-18a shows the 2D-CCF evaluated without disturbance cancellation but using a mismatched reference signal to achieve the desired sidelobe control in both the range and Doppler dimensions. In our specific case, it is not strictly necessary to apply the ASRF since the maximum range searched for detection is limited to the parking area where the tests were performed. The 2D-CCF in the Doppler dimension shows severe sidelobes structure that cannot be totally removed by using a conventional weighting network. In fact, these sidelobes represent Doppler ambiguities due to the nonregular beacon transmitting frequency. Notice that, even configuring the AP to transmit the beacon signal at 1 ms intervals, we experienced a variable beacon repetition period due to the capability of the WiFi system to “sniff” the shared medium before starting transmission. In contrast, the range sidelobes related to the Barker code due to the direct signal and other multipath contributions have been significantly reduced with respect to Figure 17-16a. Consequently, isolated peaks now appear at zero velocity at ranges of about 70 m, 100 m, and 135 m,

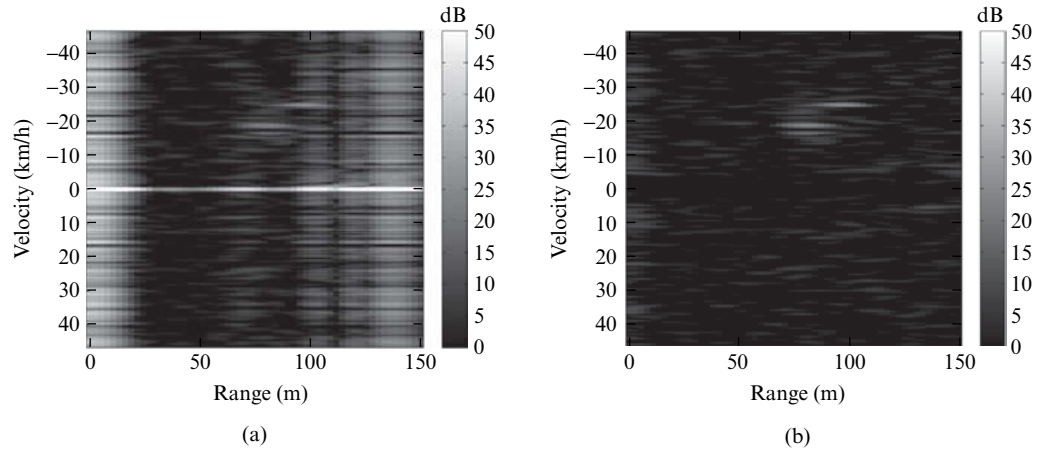


FIGURE 17-18 ■ 2D-CCF for a WiFi-based PBR with range sidelobes control. (a) Without disturbance cancellation. (b) With disturbance cancellation.

that are now clearly recognized to be the reflections of the transmitted signal over the walls, the buildings, and the metallic wire fence delimiting the parking area where the experiment was performed. The two targets returns can be identified among the residual sidelobes structures; however, their detection is still limited by the small dynamic range available. This clearly shows that the disturbance cancellation is a key stage.

When the 2D-CCF is evaluated after disturbance cancellation using a mismatched reference signal to achieve the desired ACF sidelobes reduction, Figure 17-18b is obtained. As is apparent, a better result is achieved with respect to Figure 17-16b. In that case, the human target return is significantly spread in the Doppler dimension around its actual location (range = 85 m, velocity = -20 km/h). This is in a measure due the superposition of the scattering contributions from the different body's shares moving in different manners. However, this Doppler spread of the weaker target is also related to the interaction between its return and the range-Doppler sidelobes due to the car's echo. This masking effect might be stressed for increased power level difference between the two targets echoes. When the sidelobe control networks are applied, two separate peaks appear in the map at ranges and velocities corresponding to the two targets moving in the parking area so that they can be easily detected without ambiguities. This clearly shows the significant benefit deriving from the joint application of disturbance cancellation techniques and sidelobes control filters despite the small SNR loss to be accepted. Almost identical results might be obtained when using the recovered version of the transmitted signal.

17.5.2 Range Sidelobes Control for 2D-CCF in WiMAX-Based PBR

In appropriate signals of opportunity for medium-range, high resolution passive radar surveillance we notice that wideband digital communications usually exploit an OFDM modulation primarily for its ability to cope with severe channel conditions without requiring complex equalization filters. Many wireless networking standards exploit OFDM: for example, the WiFi IEEE 802.11a and IEEE 802.11g standards as well as IEEE 802.16 standard-based technology, commonly known as WiMAX. Furthermore, the OFDM scheme has been adopted for digital video and audio broadcasting (DVB, DAB) and broadband internet access.

According to the OFDM scheme, a large number of closely spaced orthogonal subcarriers contain data divided into several parallel data streams or channels, one for each subcarrier. A cyclic prefix copying the last part of the OFDM symbol (guard interval) is inserted to prevent possible intersymbol interference. Moreover, some of the subcarriers in some of the OFDM symbols may carry pilot signals and training symbols for measuring the channel conditions and synchronization.

When an OFDM transmission is exploited by a PBR, these features must be accounted for since they are responsible for a number of undesired peaks in the signal AF that might yield a severe masking effect on weak targets or increase the false alarm rate thus significantly limiting the PBR detection capability. Obviously, the guard interval length, the number and position of pilot signals, and the training sequence are standard dependent, so their impact on the PBR performance changes with the considered application. Thus, proper techniques should be designed to cope with these undesired effects by considering the specific features of the exploited waveform of opportunity.

Wireless metropolitan area networks (MANs) are proliferating for both commercial and private use. As an example, WiMAX offers reasonable bandwidth and coverage (about 10 km) and wide accessibility [51]. Specifically, the Mobile WiMAX Standard (IEEE 802.16e) is considered in the following by exploiting the signals emitted by a base station in down-link (DL). The standard includes requirements for high data rate line-of-sight operation in the 10–66 GHz range for fixed wireless networks as well as requirements for non-line-of-sight fixed, portable, and mobile system operating in sub-11 GHz bands (2.3 GHz, 2.5 GHz, 3.3 GHz, and 3.5 GHz). Unlike many other OFDM-based systems such as WLAN, the standard supports variable bandwidth sizes between 1.25 and 20 MHz. This feature, along with the requirement for support of combined fixed and mobile usage models, makes the need for a scalable design of OFDM signal inevitable. In Table 17-6 the main scalability parameters, as recommended by the standard, are reported.

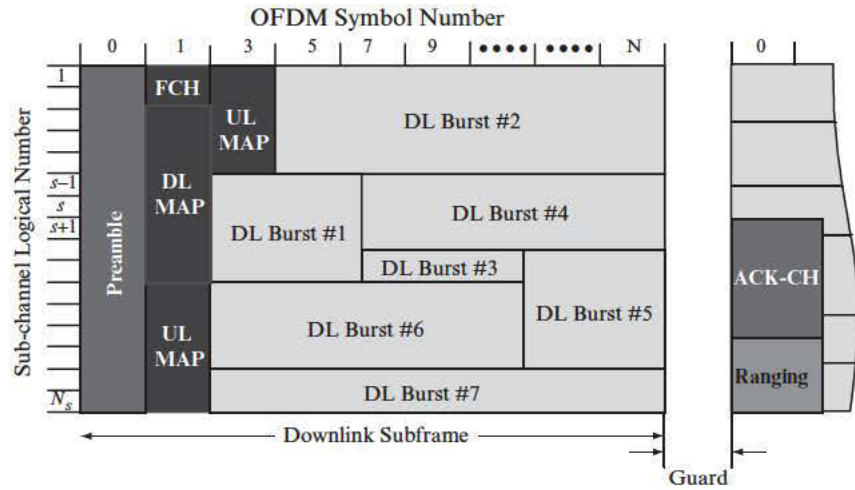
The WiMAX Standard supports a wide range of frame sizes, with durations between 2 ms (19 OFDM symbols) and 20 ms (198 OFDM symbols), to flexibly address the need for various applications and usage model requirements. Figure 17-19 shows the OFDM frame structure for time-division multiplexing mode [51]. Each frame is divided into DL and up-link (UL) subframes separated by transition gaps. Each sub-frame is subchannelized according to different subcarriers allocation modes. In DL subchannels may be intended for different receivers, whereas in UL subscriber stations may be assigned one or more subchannels and many transmitters may transmit simultaneously.

Each DL subframe starts with an OFDM symbol called preamble, employing only a subset of the available subcarriers, followed by fully used subchannelization (FUSC) or partially used subchannelization (PUSC) symbols. The preamble symbol as well as the

TABLE 17-6 ■ OFDM Scalability Parameters in WiMAX Standard

Parameters	Values				
System bandwidth (MHz)	1.25	2.5	5	10	20
Sampling frequency (MHz)	1.429	2.857	5.714	11.429	22.857
FFT size (N_{FFT})	128	256	512	1024	2048
Subcarriers spacing (Δf)	11.16071429 kHz				
Useful symbol duration (T_U)	89.6 μ s ($1/\Delta f$)				
Guard interval (T_G)	$T_U/4 = 22.4 \mu$ s; $T_U/8 = 11.2 \mu$ s; $T_U/16 = 5.6 \mu$ s; $T_U/32 = 2.8 \mu$ s				
Symbol duration ($T_S = T_U + T_G$)	112.0 μ s; 100.8 μ s; 95.2 μ s; 92.4 μ s				

FIGURE 17-19 ■
Example of OFDM
DL subframe
structure.



pilots subcarriers within FUSC/PUSC symbols are transmitted using a boosted ($4/3$ with respect to the data subcarriers) binary PSK baseband modulation with a specific pseudo noise code.

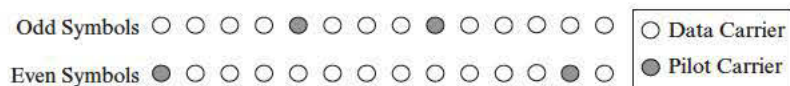
For DL FUSC, the pilot tones are allocated first and then the remaining subcarriers are divided into data subchannels. There are variable and fixed sets of pilots. The fixed sets are used in all OFDM symbols, whereas the variable sets are divided into subsets that are used in odd and even symbols alternatively. Thus, variable pilot carriers are inserted into the same positions every two OFDM symbols (i.e., a super-symbol).

For DL PUSC, the set of used subcarriers (i.e. data and pilots) is first partitioned into subchannels, and then the pilot subcarriers are allocated from within each subchannel. DL PUSC uses a cluster structure to define the pilots position, as illustrated in Figure 17-20. Each cluster spans over two OFDM symbols (in time) of 14 subcarriers, 4 of which are pilots.

For example, Figures 17-21a–b report the AF obtained for a WiMAX transmission simulated according to Standard for the $N_{FFT} = 2048$ subcarriers case and consisting of 500 DL frames with total duration of 0.5 sec exploiting a FUSC mode. The resulting bistatic radar range resolution is $\Delta R = c/B = c/(\Delta f \cdot N_{used}) = 15.76$ m, where c is the light speed, Δf is the subcarriers frequency spacing (see Table 17-6), and N_{used} is the number of actually used subcarriers (pilots subcarriers + data subcarriers + DC subcarrier).

As in the case of other waveforms of opportunity using OFDM modulation, it can be easily verified that the WiMAX signal AF contains one main peak and many side peaks. Following the approach presented in [52], these side peaks can be divided in two groups: (1) intrasymbol side peaks, generated by the correlation of the same OFDM symbol; and (2) intersymbol side peaks, generated by the correlation of different OFDM symbols. The locations of these side peaks in the delay-Doppler plane vary according to the considered transmission mode. The side peaks coordinates can be identified with reference to the DL preamble, the DL FUSC transmission, and the DL PUSC transmission modes, respectively, as described in [15].

FIGURE 17-20 ■
DL PUSC cluster
structure.



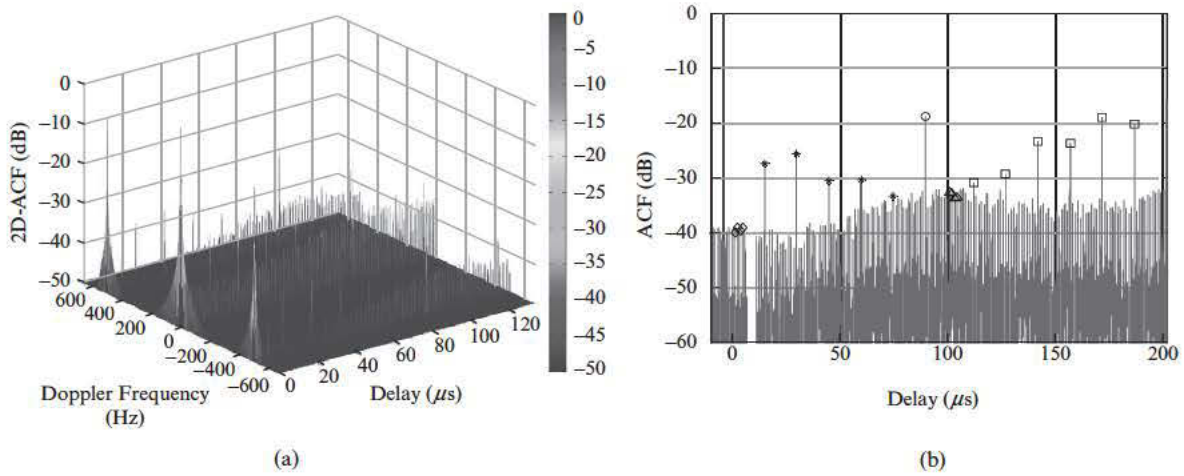


FIGURE 17-21 ■ Ambiguity function for a WIMAX DL transmissions of duration 0.5 sec exploiting a FUSC mode. (a) 3-D view. (b) Zero-Doppler cut.

For example, for a FUSC transmission mode, the following sets of coordinates have been identified:

1. Peaks caused by OFDM guard interval:

$$A_D = \left\{ \left(T_U, \frac{i}{T_S} \right); i \in \mathbf{Z} \right\} \quad (17.29)$$

where \mathbf{Z} is the set of integers and the different symbols are defined in Table 17-6. With reference to Figure 17-21b, the side peak due to guard interval, indicated with the circle, appears at $\tau = T_U = 89.6 \mu\text{s}$.

2. Peaks caused by fixed pilots.

The minimum frequency separation between fixed pilots is equal to $72 \Delta f$. Thus, the corresponding temporal periodicity is $T_U/72$. For this reason, the set A_F of intrasymbol peaks caused by fixed pilots is given by

$$A_F = \left\{ \left(\frac{nT_U}{72}, \frac{i}{T_S} \right); n \in \left[1, \dots, \left\lfloor 72 \frac{T_S}{T_U} \right\rfloor - 1 \right]; i \in \mathbf{Z} \right\} \quad (17.30)$$

In Figure 17-21b, four diamonds are used as markers for the first four intrasymbol side peaks due to fixed pilots. As expected, their relative time spacing is equal to $T_U/72 = 1.24 \mu\text{s}$.

The fixed pilots are identically distributed over consecutive transmitted symbols. Thus, the set B_F of intersymbol peaks due to fixed pilots (triangles for the first four intersymbol peaks in Figure 17-21b) is

$$B_F = \{(qT_S, 0); q \in \mathbf{N}\} \cup \left\{ \left(qT_S + T_G + \frac{nT_U}{72}, \frac{i}{T_S} \right); n \in \left[0, \dots, \left\lfloor 72 \frac{T_S}{T_U} \right\rfloor - 1 \right]; q \in \mathbf{N}; i \in \mathbf{Z} \right\} \quad (17.31)$$

where \mathbf{N} is the set of natural numbers.

3. Peaks caused by variable pilots.

As previously mentioned, in FUSC mode the variable pilots position is different for even- and odd-numbered OFDM symbols. However, frequency spacing among variable pilots within an OFDM symbol is constant and equal to $12\Delta f$. Considering the alternate transmissions of odd and even symbols and the relative displacement of the pilots subcarriers sets between consecutive symbols, the whole set A_V of intrasymbol side peaks due to variable pilots can be written as

$$A_V = \left\{ \left(n \frac{T_U}{6}, \frac{i}{T_S} \right); n \in \left[1, \dots, \left\lfloor 6 \frac{T_S}{T_U} \right\rfloor - 1 \right]; i \in \mathbf{Z} \right\} \cup \left\{ \left((2n+1) \frac{T_U}{12}, \frac{2i+1}{2T_S} \right); n \in \left[1, \dots, \left\lfloor 6 \frac{T_S}{T_U} \right\rfloor - 1 \right]; i \in \mathbf{Z} \right\} \quad (17.32)$$

Notice that the temporal separation, $T_U/6$, between peaks corresponds to the inverse of the frequency displacement, $6\Delta f$, between the pilots subcarriers sets exploited by odd and even symbols. In Figure 17-21b, the first five intrasymbol side peaks due to variable pilots are labeled with stars.

Similarly the set B_V of intersymbol side peaks caused by variable pilots is given by

$$B_V = \{(2qT_S, 0); q \in \mathbf{N}\} \cup \left\{ \left(qT_S + T_G + n \frac{T_U}{6}, \frac{i}{T_S} \right); n \in \left[1, \dots, \left\lfloor 6 \frac{T_S}{T_U} \right\rfloor - 1 \right]; q \in \mathbf{N}; i \in \mathbf{Z} \right\} \cup \left\{ \left(qT_S + T_G + (2n+1) \frac{T_U}{12}, \frac{2i+1}{2T_S} \right); n \in \left[1, \dots, \left\lfloor 6 \frac{T_S}{T_U} \right\rfloor - 1 \right]; q \in \mathbf{N}; i \in \mathbf{Z} \right\} \quad (17.33)$$

In Figure 17-21b, the first six intersymbol side peaks due to variable pilots are labeled with squares. As expected, the first peak of this set appears at $\tau = T_S + T_G + T_U/6 = 126.9 \mu s$ ($T_G = T_U/8$ in the considered example). Notice that the level of side peaks due to variable pilots is greater than the level of side peaks due to fixed pilots since a higher number of subcarriers is devoted to variable pilots.

Obviously, side peak distributions are different for FUSC and PUSC transmission modes [15]. However, the obtained PSLR is comparable in the two cases. Specifically, the PSLR is dominated by the peak caused by the guard interval whose level has a theoretical value equal to $20\log_{10}(T_G/T_U) = -18$ dB with respect to the main peak.

Regarding the distribution of side peaks in the Doppler dimension, the first Doppler side peak set (due to the preamble subcarriers allocation) appears at a Doppler frequency corresponding to the inverse of the temporal separation between consecutive DL frames ($1/T_{frame}$). However, the preamble contribution to the AF of the whole DL frame can be reasonably neglected. In any case, the single OFDM symbol corresponding to the preamble might be left out of the integration time with a very small SNR degradation. Consequently, the first Doppler side peak set due to pilot tones appears at a Doppler frequency equal to $f_D = 1/2T_S = 4.96$ kHz. This Doppler frequency value corresponds to a bistatic velocity of about 765 km/h (in the worst case a 3.5 GHz carrier frequency). However, additional sidelobe structures appear at a Doppler frequency equal to 500 Hz (Figure 17-21). Such peaks are obviously because of the pulsed nature of the considered signals that consist of DL frames of duration T_{DL} transmitted every T_{frame} . The Doppler ambiguities result from the equivalent PRF $= 1/T_{frame}$, which is not within the control of the radar designer. As an example, for the 2 ms frame duration considered in Figure 17-21,

the first Doppler ambiguity appears at about 500 Hz corresponding to a bistatic velocity of 77 km/h.

First, we focus our attention on removing the side peaks appearing at zero Doppler. To this purpose we might use the approach presented in [10] that was designed for DVBT-based PBR exploiting an OFDM modulation. Accordingly, the unwanted peaks removal is performed by preprocessing the reference signal with a linear filter based on the knowledge of the expected value of the WiMAX OFDM signal ACF. The theoretical ACF can be evaluated using the approach in [16] where the expected value of the ACF has been derived for OFDM WiFi signals. The unpredictable component of the considered waveform of opportunity is mainly related to the data subcarriers; thus, averaging over this component usefully characterizes the deterministic components responsible for the regular structures in the AF. In practical applications, long integration times are usually required to attain a certain level of performance (SNR, Doppler resolution). Notice that increasing the integration time results in an average AF that resembles its statistical expectation.

In particular, we observe that the expected value of the ACF is a periodic function with period equal to the symbol duration (for PUSC case) or two times the symbol duration (for FUSC case). This implies that an inverse filter can be designed that is able to whiten the average waveform spectrum. The filter coefficients have a support with the length of a single period duration. By solving a proper inverse filter problem, we obtain

$$w = FFT \left(\frac{1}{FFT(\alpha[k])} \right) \quad (17.34)$$

where $\alpha[k]$ is the expected ACF. As is apparent, the filter can be easily applied in the frequency domain by Fourier transforming the input data and then performing the inverse FFT on the filtered data. When computing the filter weights, only the ACF values at the peaks locations might be considered to build up the system of linear equations required to obtain the zero-Doppler peaks suppression. This results in a reduced computational load and in a less noisy weights evaluation since the average ACF is almost identical to its statistical expectation at the peaks locations. In addition, the filter weights must be computed just once. Figures 17-22a–b show the mismatched ACF obtained by cross-correlating

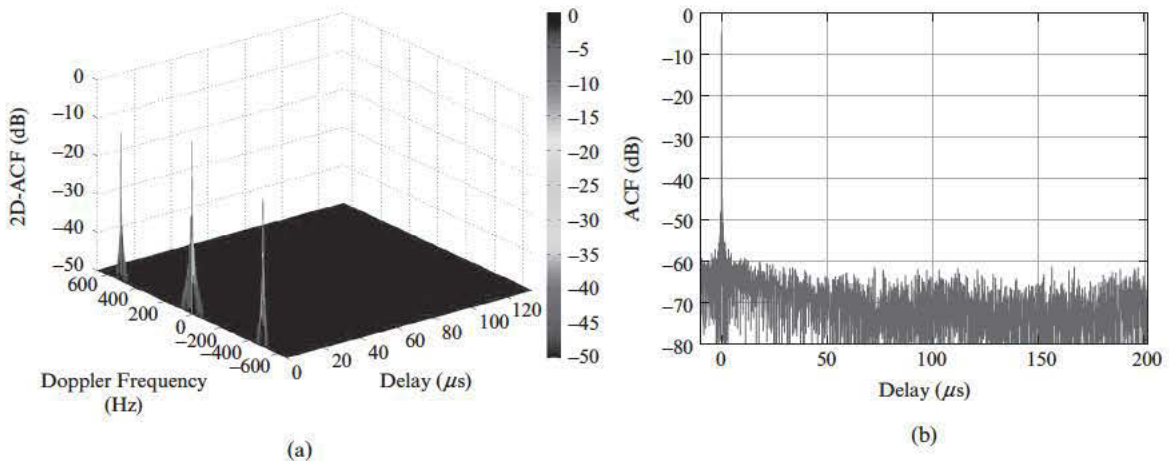


FIGURE 17-22 ■ Ambiguity function after peaks removal for a FUSC mode. (a) 2-D view. (b) Zero-Doppler cut.

the original WiMAX signal with its prefiltered version after the application of the proposed peaks removal approach for the same case of Figure 17-21. A conventional weighting network also has been applied in the frequency domain to improve the PSLR resulting from the rectangular shaped power spectral density (PSD) of the OFDM signal. The considered approach is able to effectively remove the undesired deterministic peaks due to the presence of guard interval and pilot subcarriers in the OFDM symbols, with only a very limited SNR loss.

The side peaks in the Doppler dimension resulting from the temporal separation between the integrated pulses (i.e., a low equivalent PRF) cannot be counteracted and yield Doppler ambiguities. Such ambiguities become less severe when decreasing the frame duration as foreseen for WiMAX. However, after applying the proposed filter, only the zero-delay peaks at each ambiguous Doppler frequency survive, whereas the side peak set previously present at positive delays is correctly removed. The reported results refer to particular selections of the WiMAX standard parameters according to the available technical literature; since the WiMAX standard definition has not been completed yet, additional transmission modes could be designed in the future requiring a dedicated analysis. However, the considered analysis approach and the proposed filtering strategy will be still valid for any transmission mode exploiting an OFDM modulation. As an example, the application of this approach to WiFi signals transmitted using an OFDM scheme is presented in [21]. In the next subsection its application to DVB-T signals is considered.

Finally, when considering practical applications of a WiMAX-based PBR, additional issues should be taken in account. Since a dynamic bandwidth allocation is adopted, according to the traffic in the network, a corresponding performance degradation should be expected at least in terms of range resolution. Moreover, a base station might be transmitting a great variety of signals in FUSC, PUSC, or optional modalities; consequently, average AF characteristic should be expected among the different transmission modes considered in this subsection, and proper processing strategies should be designed to correctly manage mixed transmissions.

17.5.3 2D-CCF Shape Control for DVB-T Signals

With reference to long-range passive radar surveillance, the linear filter for sidelobe reduction presented in the previous subsection is now applied to DVB OFDM signals. Alternative nonlinear approaches are also described to cope with the masking effect due to undesired peaks appearing in the AF of such signals.

The DVB-T standard is proliferating and rapidly replacing its analog counterpart [53]. Many countries have already switched or planned to switch to the DVB-T standard. A number of studies have looked at the possibility of using the DVB-T transmissions as waveform of opportunity for PBR systems [6–10]. These signals show both good coverage and wide frequency bandwidth and result in a reasonably long-range surveillance capability with remarkable range resolution.

The DVB-T signal exploits an OFDM modulation whose main parameters are listed in Table 17-7 for typical 8 MHz channels, distinguishing between the $2k$ mode and $8k$ mode specifications. The data stream is modulated over a number of evenly spaced orthogonal subcarrier frequencies, contemporaneously transmitted for a time interval equal to T_U . A guard interval of duration T_G is inserted by cyclically extending each OFDM symbol so the total symbol duration is equal to $T_S = T_U + T_G$. The subcarriers are separated by $\Delta f = 1/T_U$ and the number of sub-carriers depends upon the considered mode. However,

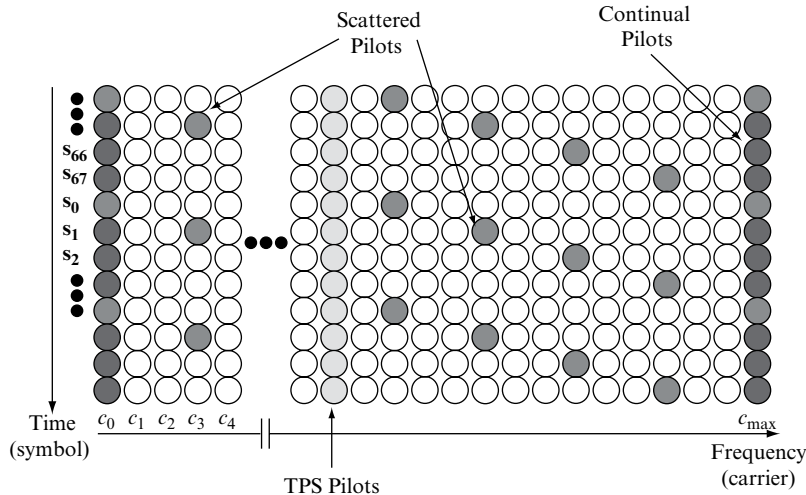
TABLE 17-7 ■ Main Parameters of 2k and 8k Modes DVB-T Signals

Parameter		2k mode	8k mode
Number of subcarriers		1705	6817
Useful symbol duration T_U		224 μs	896 μs
	$T_U/32$	7 μs	28 μs
	$T_U/16$	14 μs	56 μs
Guard interval T_G	$T_U/8$	28 μs	112 μs
	$T_U/4$	56 μs	224 μs
Bandwidth		7.61 MHz	7.61 MHz

the above parameters are scaled so that the subcarriers span the 7.61 MHz bandwidth allocated to the DVB-T channel. The DVB-T signal structure is organized in frames with each frame consisting of 68 OFDM symbols, as shown in Figure 17-23. In addition to the data carriers, the DVB-T signal contains pilot carriers and transport parameter signalling (TPS) carriers, which are used for receiver synchronisation, channel estimation, and system parameters transmission. To this aim, the pilots carriers are transmitted at boosted power level $E_p = 16/9$ while the power level of data carriers and TPS pilots is normalized to 1. The pilot and TPS carriers are scattered at given frequencies, and according to a known rule [53]. In particular, the pilots carriers include continual pilots having a fixed frequency position from symbol to symbol and scattered pilots, whose frequency positions are changed from symbol to symbol over four symbols period (i.e. a super-symbol).

For example, Figure 17-24 reports the AF for a simulated DVB-T signal of duration 20 ms in 2k mode for $T_G = T_U/32$ and shows the number of side peaks due to the guard interval and the pilot carriers. Specifically, in 2k mode, the peak generated by guard interval occurs at $\tau = 224 \mu\text{s}$, whereas the peaks due to pilot subcarriers can be divided into intrasymbol peaks ($0 \leq \tau \leq T_S$), and intersymbol peaks ($\tau > T_S$).

As for the WiMAX signal, the position of the undesired peaks in the range-Doppler plane can be identified based on the parameters adopted by the standard. Similarly, the expected value of the AF can be evaluated. Thus, the linear filter for range sidelobes reduction can be modified according to (17.34) by using the actual expression for the signal ACF. The results obtained for the signal considered in Figure 17-24 are reported in Figure 17-25.

**FIGURE 17-23** ■ DVB-T signal frame structure.

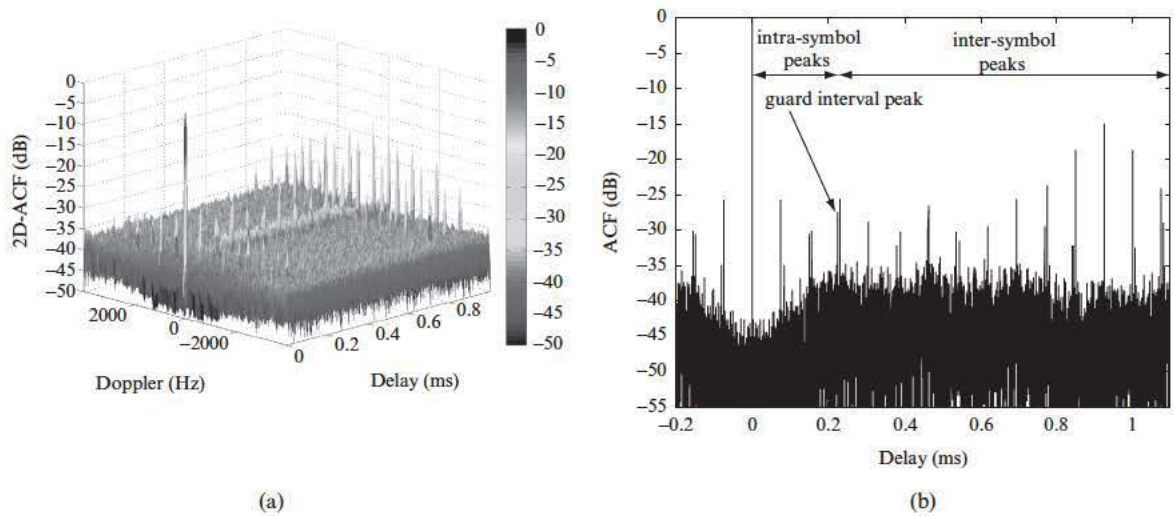


FIGURE 17-24 ■ Ambiguity function for a DVB-T signal of duration 20 msec in 2k mode. (a) 3-D view. (b) Zero-Doppler cut.

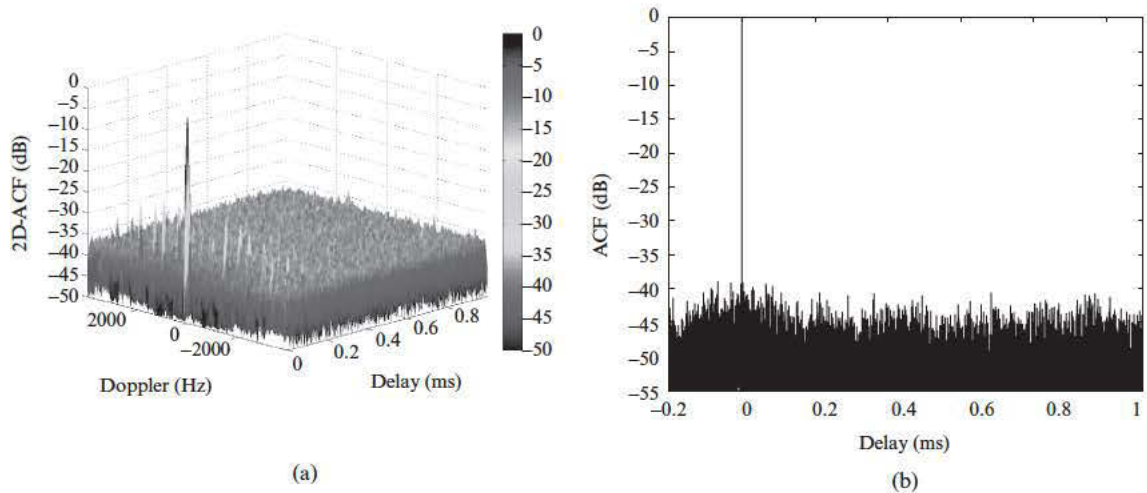


FIGURE 17-25 ■ Ambiguity function after sidelobe control using the linear filter for a DVB-T signal of duration 20 ms in 2k mode. (a) 3-D view. (b) Zero-Doppler cut.

As is apparent, the described approach is able to mitigate the undesired deterministic peaks due to the presence of the guard interval and of the pilot carriers in the OFDM symbols. Specifically, the unwanted peaks at zero Doppler have been correctly removed (see Figure 17-25b); however, some peaks are left at different Doppler frequencies.

A removal of the undesired peaks over the whole range-Doppler plane can be achieved by exploiting the non linear approaches presented in [9, 52]. They have a common scheme and are summarized in the block diagram sketched in Figure 17-26.

These approaches exploit the knowledge of pilot carrier positions inside the OFDM symbols for the removal of the unwanted deterministic peaks in the DVB-T signal AF. In particular, continual pilots have fixed positions from symbol to symbol, while scattered pilot carriers are inserted into the same positions every four OFDM symbols (i.e., the

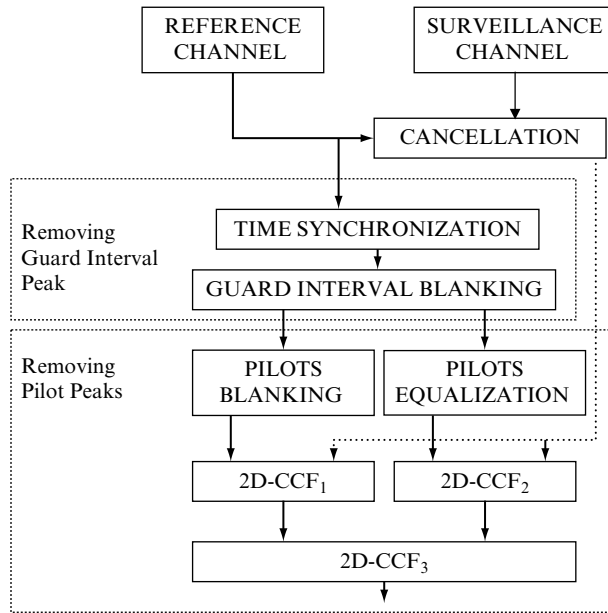


FIGURE 17-26 ■ Block diagram of the DVB-T signal AF improvement technique proposed in [9,52].

super-symbol). Therefore, time synchronization is first required as a necessary operation. After the frame synchronization of the reference signal, the guard interval blanking removes the guard interval peak. Then, the intrasymbol and intersymbol peaks are removed through the pilot equalization and the pilot blanking, respectively. These two procedures are performed in two parallel stages, because blanking and equalization cannot be applied simultaneously. Two different 2D-CCFs are evaluated ($2D\text{-CCF}_1$ and $2D\text{-CCF}_2$ in Figure 17-26), as a result of the correlation between the surveillance signal (after disturbance cancellation) and the two prefiltered versions of the reference signal. Finally, $2D\text{-CCF}_3$ is obtained by nonlinearly combining $2D\text{-CCF}_1$ and $2D\text{-CCF}_2$ (e.g., minimum operation).

The results obtained at the different stages for the same case of Figure 17-25 are reported in Figure 17-27. In Figure 17-27a, after pilot blanking, the intersymbol peaks are removed, and intrasymbol peaks are left. In contrast, in Figure 17-27b, after pilot equalization the intrasymbol peaks are removed and intersymbol peaks are left. By combining the two results, all the unwanted peaks are correctly removed.

The linear filter in [10] is fully effective for zero-Doppler shift, whereas the nonlinear approaches in [9, 52] theoretically provide the removal of undesired peaks at Doppler frequencies other than zero. However, other considerations should be taken into account when comparing their performance. While the SNR loss obtained with the different approaches is largely comparable, the linear approach does not require two complementary and parallel processing stages, thus reducing the overall computational load and the system complexity. Furthermore, as a consequence of the parallel processing and the following nonlinear combination of two different 2D-CCF, the nonlinear approaches proposed in [9, 52] have been shown to yield undesired residual peaks in some particular scenarios, when more than one target is present, located at different time delays [10]. In fact, although the nonlinear minimum combination works extremely well for a single target, it shows a potential weakness when multiple or extended targets are present. Specifically, pilots equalization yields intersymbol spurious sidelobes, and pilots blanking yields intrasymbol spurious sidelobes (see Figures 17-27a–b). When the spurious peaks are temporally

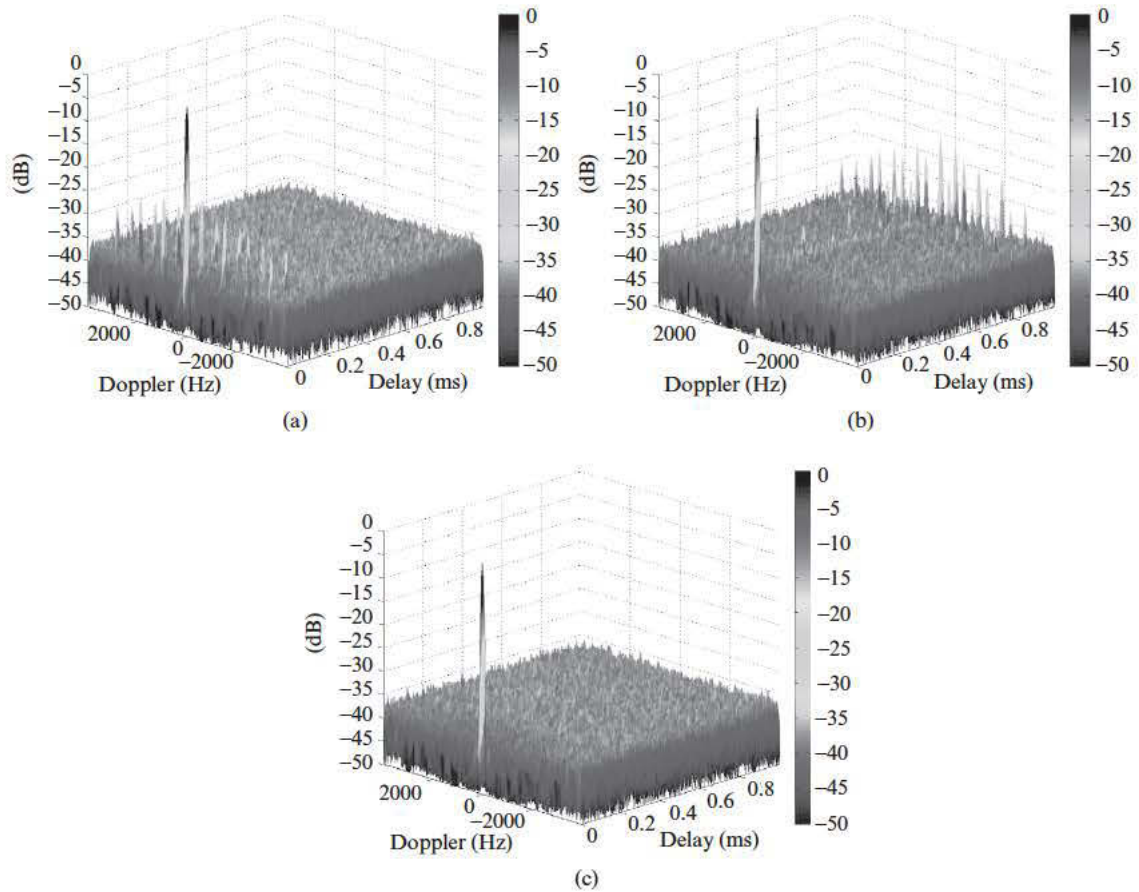


FIGURE 17-27 ■ Ambiguity function after sidelobe control using the nonlinear approach in [9,52] for a DVB-T signal of duration 20 msec in $2k$ mode. (a) $2D\text{-}CCF_1$ (after pilot blanking). (b) $2D\text{-}CCF_2$ (after pilot equalization). (c) $2D\text{-}CCF_3$ combining (a) and (b).

displaced, the nonlinear combination (minimum operation) achieves the desired result. However, when the spurious peaks are colocated, the combination is no longer able to remove them and false targets (ghosts) appear in the scene.

The results presented in this subsection have been obtained for a small guard interval duration (i.e., $T_G = T_U/32$). The guard interval blanking required by the nonlinear techniques yields high peaks in the corresponding signal AF appearing at Doppler values multiples of $(T_G + T_U)^{-1} = 4329$ Hz in the $2k$ mode. These peaks are not visible in Figure 17-27 where the Doppler axis has been limited to 4 kHz, which certainly includes the typical values for the targets of interest. However, when larger guard intervals are exploited, especially in mode $8k$ (see Table 17-7), these peaks would appear at Doppler frequencies of practical interest.

In addition, the nonlinear approaches intrinsically require time and frequency synchronization. In contrast, in the linear approach the synchronization is avoided based on the observation that, when using an integration time equal to an integer number of super-symbols, the expected value of the AF has a behavior independent on the starting time. Based on this property, an appropriate filter can be designed to remove all sidelobes of the AF. In addition, if the expected value of the AF is obtained based on a sample

estimation of the AF in multiple previous time intervals, also frequency offsets and other synchronization errors are intrinsically removed.

The effects of synchronization errors on the different techniques have been analyzed in [54] with reference to synchronization error, carrier frequency offset, and sampling frequency offset. Specifically, time synchronization errors are shown to yield a strong performance degradation in terms of PSLR achievable with the nonlinear approaches; the performance degradation starts for time synchronization error lower than the guard interval duration T_G . Similarly, the PSLR obtained with the nonlinear approaches rapidly decreases also for low frequency offset values; when the frequency offset is equal or higher than $1/T_U$ (subcarrier spacing) (i.e., integer carrier frequency offset) PSLR decreases to its lower bound, represented by the PSLR of the standard DVB-T AF. In contrast, the PSLR obtained with the linear filter approach is independent of time synchronization and frequency offset errors. Nevertheless, it shows a significant performance degradation when a sampling frequency offset higher than 3 ppm is present, whereas the nonlinear methods appear to be more robust. However, this degradation is inessential in most practical cases of interest, because typical sampling error values are lower or equal to ± 1 ppm.

17.6 | MULTICHANNEL PROCESSING FOR DETECTION PERFORMANCE IMPROVEMENT

In this section we show that multichannel processing has significant potential to improve PBR performance. Multichannel processing can be obtained by exploiting parallel channels that differ by frequency, polarization, time, and space. Although there is not space here to comprehensively treat all possibilities, we show how the detection performance of PBR can be improved by jointly exploiting multiple receiving channels with reference to a selection of three multichannel types for an FM-based PBR:

- Multiple frequencies. With reference to multiple radio channels, we explore different ways to combine them and to control the adaptive detection threshold. Performances are evaluated against live data.
- Multiple polarizations. For this case we consider how to exploit polarizations to increase detection rate. The experimental results are shown again for FM radio-based PBR.
- A receiving antenna array. For this case we consider the signals received at the multiple receiving elements of the antenna array and show the potential approaches to exploit them to (a) reduce the direct signal contribution and (b) steer the receiving beam on the desired target region. Some hints are given for the array configuration optimization.

17.6.1 Linear–Nonlinear Integration of Multiple Frequency Passive Radar Channels

Despite the effectiveness of the processing techniques introduced in the previous sections, the PBR detection and localization performance is time varying and largely depends on the following:

- The instantaneous characteristics of the transmitted waveform
- The instantaneous characteristics of the electromagnetic environment (i.e., the radio propagation channel)

However, in many PBR applications, we might assume that the exploited transmitter of opportunity is simultaneously broadcasting different signals at different carrier frequencies within the assigned band. The availability of N_{ch} different frequency channels provides a large diversity, and it is likely that the joint exploitation of the corresponding signals yields a significant performance improvement for the resulting multiple-frequency (MF) PBR. Specifically, the signals received at different frequency channels might be jointly exploited to:

1. Make the detection scheme robust with respect to both the content of the broadcast channel and the propagation channel conditions, thus improving the detection performance of the PBR and its reliability.
2. Improve the localization capability of the system by properly combining the measurements obtained using the signals transmitted at different carrier frequencies.

The first issue is mainly addressed in the following. PBR systems exploiting FM radio signals are discussed since they are very well suited for the proposed MF approach. In fact, an FM radio transmitter usually broadcasts different radio channels in the 88–108 MHz band that can be simultaneously acquired by a single receiving channel with reasonable bandwidth. Moreover, the FM-based PBR represents a worse case since it has been recognized to be highly affected by the time-varying characteristics of both the transmitted waveform and the propagation channel conditions [47, 55–58], as briefly described in the following.

The standard assigns a nominal frequency bandwidth of 200 kHz to each FM radio broadcast channel. However, the PSD of a FM signal might experience a considerable variability with dependence on the program content and the time of transmission. For example, an FM speech channel shows a high degree of temporal variability in PSD width and in turn in range resolution; in addition, the presence of long periods of silence or speech pauses can dramatically degrade the achievable range resolution. In contrast, music programs usually lead to wider PSD and exhibit less variation, thus resulting in a better range resolution, [3, 47]. Therefore, the actual range resolution is highly varying with the instantaneous transmitted waveform and thus with the considered FM radio channel. Figure 17-28 reports the spectrograms of different data files (of about 1.1 sec) collected at different times and FM radio channels. It clearly shows the fast temporal variability of the FM signal instantaneous frequency bandwidth and the highly different behavior of the waveforms contemporaneously transmitted at different carriers.

A second highly varying characteristic of the AF concerns the level of its sidelobes. Figure 17-29 shows the PSLR evaluated for 32 different real data tracks with different program contents. With the exception of the silence leading to a periodic AF caused by the stereo pilot tone (totally ambiguous, thus yielding a PSLR of 0 dB), a PSLR of about 20 dB is typical for music tracks. The low values confirm that significant masking effects can be present for weak targets. It is likely that a voice track contains silence pauses, so it yields usually slightly worse PSLR. The three categories of content (music, voice, silence) are typically all present in turn on each FM channel, with a mixing largely depending on the type of radio channel with a large variability inside the FM band.

Recall that the global power transmission level of the considered FM radio channels strongly affects the PBR detection performance. The transmitting stations broadcast many radio programs on different FM channels, with different power levels due to different coverage requirements. An FM radio channel can experience an abrupt power loss due

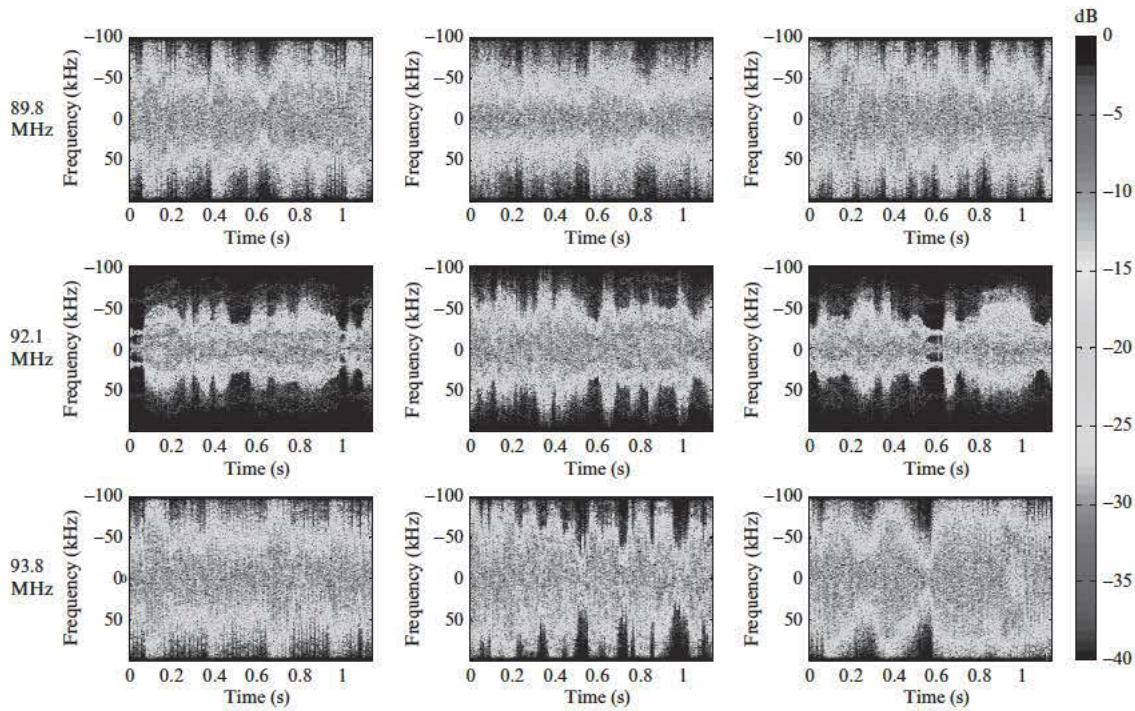


FIGURE 17-28 ■ Spectrograms of the signals collected at different times for different FM radio channels.

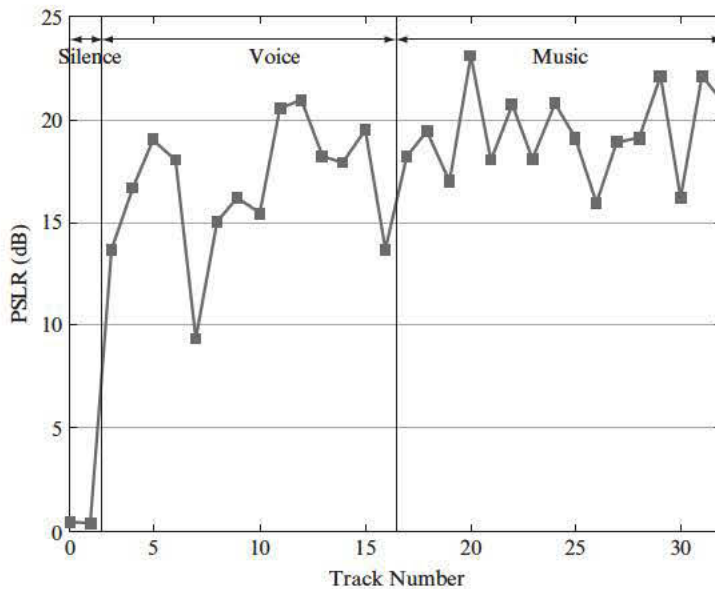


FIGURE 17-29 ■ PSLR of the waveform ambiguity function evaluated over different real data tracks.

to a temporary failure, a coverage planning modification, or just a change of coverage between different hours of the day time. These effects should be taken into account when a continuous reliability is to be guaranteed for the resulting PBR.

Among the characteristics of the environment, the multipath reflections can significantly affect the PBR performance. Their effect on both the reference signal and the

surveillance signal is time varying and can depend on the specific transmission carrier frequency, the meteorological conditions, and the soil moisture at the reflection point. Thus, even when receiving from the same transmitter, the propagation channel conditions at different FM radio channels might be different.

Another typical effect of the electromagnetic environment on the PBR is the presence of co- and interchannel interferences. Due to the high density of FM radio transmissions in the 88–108 MHz band (and despite the regulations present in each country), it is likely that (1) the same frequency is used by different transmitters covering different but neighboring geographical areas and (2) a high power transmission is present on a frequency channel adjacent to the selected channel. These undesired signals act as interferences for the selected channel bandwidth, thus increasing the system noise floor and reducing the achievable detection performance. This especially applies in rural areas covered by many high-power transmitters. The power ratio between co-/adjacent-channel transmissions (usually prescribed at about 20/30 dB) might be insufficient since often the surveillance antenna is pointed toward an opposite direction with respect to the exploited transmitter of opportunity, thus enhancing the level of the signals received from other transmitters at the same/adjacent channels.

As is apparent, for a selected FM radio channel, the time-varying characteristics of both transmitted waveform and electromagnetic environment yield PBR performance that changes with time. This behavior might strongly affect the instantaneous reliability of a PBR system. Even choosing a good channel (namely, showing in the average high-power level, range resolution and PSLR, together with a nice propagation channel), the instantaneous PBR performance might experience large degradations. These considerations suggest that the use of a single-frequency channel (SFC) does not always guarantee a high-quality PBR detection.

However, it is quite reasonable to assume that, while the instantaneous characteristics of a specific FM radio channel are highly varying in time, for any assigned time instant it is possible to identify one or more FM radio channels yielding signals with desirable characteristics for PBR purposes. To take advantage of this situation, the PBR system should be able to collect at the same time the signals at different carrier frequencies and appropriate processing strategies must be designed to exploit the large diversity of information conveyed by different FM radio channels.

To jointly exploit N_{ch} FM radio channels, we assume that they are transmitted by the same transmitter and received by the same PBR receiver, so a specific target is observed at all channels with the same bistatic range and velocity. However, different radio channels have different carrier frequencies so that the echo from the same target is received with different Doppler frequencies. Clearly, if the same integration time T_{int} is used at each carrier frequency f_n , the same Doppler resolution is achieved, that in turn yields a slightly different velocity resolution Δv_n . To obtain range-velocity maps at the different FM radio channels that are directly comparable, slightly different integration times T_{intn} are used on the N_{ch} channels according to the criterion (with c = light speed):

$$\Delta v_n = \frac{c}{f_n T_{intn}} = \Delta v \Rightarrow T_{intn} = \frac{c}{\Delta v \cdot f_n} \quad n = 0, \dots, N_{ch} - 1 \quad (17.35)$$

In this way the range-velocity maps of the different FM channels can be integrated pixel by pixel, and the coincidences of the detections can be exploited both to increase the detection capability (reducing the undesired degradations of the single channel waveform/environment) and to reduce false alarms.

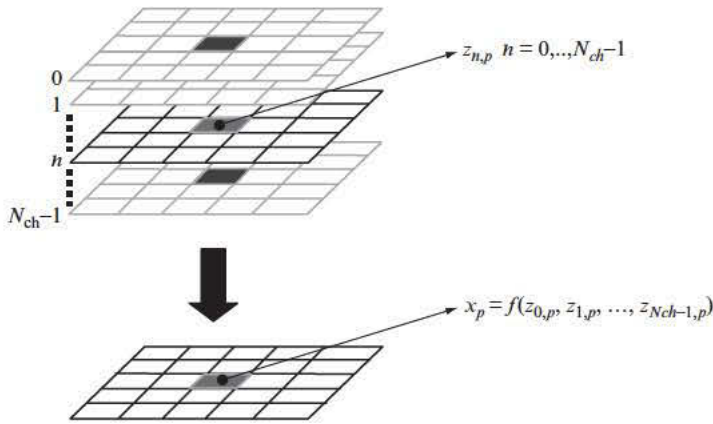


FIGURE 17-30 ■ Sketch of the MF approach.

To this purpose we consider a noncoherent integration. The potentialities of noncoherent integration are well-known in radar literature and used to increase the detection performance by exploiting the target echoes received on multiple channels and to average over the fluctuation of the target radar cross section (especially when using frequency agility). However, for the MF PBR, we aim at exploiting the noncoherent integration to make the detection scheme robust with respect to the time-varying characteristic of the received waveform. Therefore, together with the standard integration approaches, we look for specific integration strategies aiming at mitigating the drawbacks caused by the random waveform used. The following approaches are considered [55–58].

Centralized Linear Integration (SUM) Among the standard integration approaches that operate with the full dynamic range, the linear integration (after square detector) is largely the most common solution. The value x_p of the p -th pixel in the map after integration can be written starting from the values of the corresponding pixels $z_{n,p}$, $n = 0, \dots, N_{ch} - 1$ at the N_{ch} maps (see Figure 17-30) as

$$x_p = \sum_{n=0}^{N_{ch}-1} z_{n,p} \quad (17.36)$$

Assuming that most of the exploited FM channels yield a reasonable SNR at the range-velocity location of a given target, their 2D-CCFs are incoherently summed aiming at enhancing the resulting SNR on the integrated map. With this approach, the instantaneous characteristics of the bad channels are likely to be averaged with the good ones, thus yielding a limited impact on the final detection performance.

Decentralized Integration L/N_{ch} (DEC) The alternative approach, usual for noncoherent integration with standard radar systems, is to use a binary integration after applying a first detection threshold separately on each channel. According to this strategy, a detection is declared at a given pixel when L detections of N_{ch} channels are obtained for the considered pixel over the SFC maps. This is expected to yield reasonable detection performance improvement when at least L of the N_{ch} exploited FM radio channels yield a reasonable SNR at the range-velocity location of a given target. The advantage of the nonlinear integration rule in the presence of one or few ($< N_{ch} - L$) FM radio channels with a bad waveform with high sidelobes is related to its capability to remove the corresponding effect on the final detection.

In addition to the standard integration approaches already outlined, it is interesting to attempt the exploitation of nonlinear detection rules able to remove the impact of the bad waveforms with high sidelobes, together with an integration approach operating with the full dynamic range of power values (also known as centralized strategy). In particular, the maximum and the minimum modulus can, respectively, be considered. The former guarantees that the highest value is always retained to avoid degrading good target echoes, whereas the latter aims at considering only very reliable targets, whose echoes are received with a reasonably high value at all FM radio channels.

Centralized Nonlinear Integration MAX Assuming that only one or few of the exploited FM radio channels yield a significant SNR at the range-velocity location of a given target, a maximum approach should be used by selecting, for each location, the maximum value among the corresponding SFC maps locations. In this case, the pixel value after integration can be written as

$$x_p = \max \{z_{n,p}\}_{n=0,\dots,N_{ch}-1} \quad (17.37)$$

This approach has the drawback of preserving the sidelobes structures due to the instantaneous characteristics of a single FM channel.

Centralized Nonlinear Integration MIN Aiming at significantly reducing the high peaks in the 2D-CCF due to the instantaneous characteristics of the single FM radio channel, a minimum approach should be implemented by selecting, for each map location, the minimum value among the corresponding SFC maps locations. In this case, the pixel value after integration can be written as

$$x_p = \min \{z_{n,p}\}_{n=0,\dots,N_{ch}-1} \quad (17.38)$$

This approach is expected to yield reasonable detection performance only when all the exploited FM radio channels yield a reasonable SNR at the range-velocity location of a given target.

To allow the practical use of the MF integration strategies, proper detection schemes must be designed including the control of the false alarms over the integrated map. For a decentralized integration strategy, a CFAR threshold is separately applied at the single channel map and the detection results are subsequently combined. Using a centralized strategy, the three different integration approaches are considered to build up the final integrated map (see Figure 17-30). In this case, before the different range-velocity maps are integrated, a proper normalization should be applied. To this purpose, the actual noise floor can be estimated over a map portion located at high range values, where the target contributions and the effects of the waveforms (lobes main structures) can be assumed to be negligible. This value is used to scale each map to yield the same final noise floor. A standard detection scheme can be then applied to obtain the desired CFAR condition. As an example, using a CA-CFAR, the cell under test x_0 is compared with an adaptive threshold, obtained by scaling by a factor G_{CA} a local estimate x_{CA} of the residual noise intensity level based on P secondary data x_1, \dots, x_P :

$$x_0 \underset{H_0}{\overset{H_1}{>}} G_{CA} \cdot x_{CA} = \frac{G_{CA}}{P} \sum_{p=1}^P x_p = \frac{G_{CA}}{P} \sum_{n=0}^{N_{ch}-1} \sum_{p=1}^P z_{n,p} \quad (17.39)$$



FIGURE 17-31 ■ Sketch of the acquisition geometry.

Obviously, the scaling factor G_{CA} should be modified, taking into account the exploited integration strategy over the N_{ch} frequency channels [57].

A comparison of the proposed integration strategies is reported in the following in terms of achievable detection performance against real data. We considered the data collected by a wideband PBR prototype developed and fielded at the DIET Department of the University of Rome “La Sapienza” [55–58], based on a direct radio frequency sampling approach. We show the results out of the acquisition campaigns that were carried out on May 17 and May 18, 2007, by Riva di Traiano (Civitavecchia) about 70 km northwest of Rome, Italy. The reference antenna was steered toward a transmitter located about 70 km away, at Monte Argentario (approximately pointing to northwest), as illustrated in Figure 17-31. The surveillance antenna was pointed at about 180° (approximately southeast) and covers, in its 3 dB pattern, the standard departure routes of the Leonardo da Vinci Airport in Fiumicino (basically most of the routes for the northwest of Europe) as well as a number of the standard arrival routes that lead the aircrafts to join the descend path of the ILS for runway 34. A detailed analysis of the data in both temporal and spectral domain suggested the selection of a limited number of FM radio channels (SFC 1–5) inside the available frequency bandwidth (see Table 17-8). During the considered acquisition campaigns, several sequential data acquisitions were performed; each acquisition is of about 1.1 sec duration, with a temporal spacing of about 23 sec between two consecutive acquisitions. Data sets acquired with this approach cover a total time duration of about 1 hour (approximately 150 data files) for each day of the acquisition campaigns. Live Air Traffic Control (ATC) registrations were also collected for the same air space.

To validate the performance of the SFC operation, all the available data files have been processed, according to the processing techniques described in the previous sections. For example, Figure 17-32 shows the sequences of the detection results obtained for three consecutive data files on the same range-velocity map, separately for some of the selected FM radio channels. The available ATC registrations have also been reported in each figure (small plots) for comparison. Different markers are used to indicate the plots detected by the PBR at different file times. The identified plot sequences are indicated with circles using

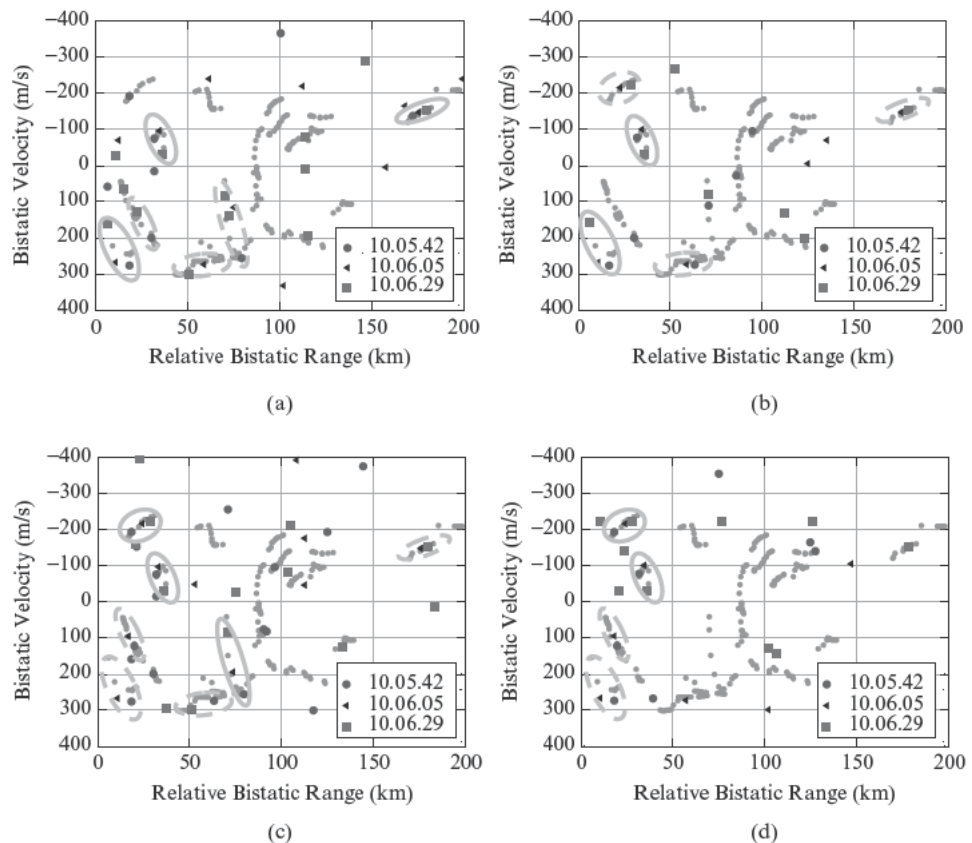
TABLE 17-8 ■ Selected FM Radio Channels Information

ID	Frequency (MHz)	FM Radio Broadcast	Polarization	DNR Ref Channel (dB)	DNR Surv Channel (dB)
SFC1	89.8	M20	Vertical	50.5	38.6
SFC2	92.1	RAI R.2	Horizontal	56.0	35.6
SFC3	93.8	Radio Dim. Suono Roma	Vertical	52.8	34.7
SFC4	94.3	RAI R.3	Horizontal	57.1	41.8
SFC5	91.3	Lattemiele	Vertical	51.3	35.6

the continuous line for complete sequences containing three plots and the discontinuous for sequences containing only two plots.

On some of the considered channels, quite complete target plot sequences are observed, whereas on others only a few target plots are detected. Furthermore, the exploitation of different channels allows different target plots within the same sequence to be detected. Missed detections are reasonably due to general degradation of the characteristics of both the transmitted waveform and the electromagnetic environment. However, the FM radio channels at 89.8 MHz (SFC1) and 93.8 MHz (SFC3) generally yield better detection performance due to reduced co- and interchannel interferences that set the average performance level.

FIGURE 17-32 ■ Plot sequences for different FM radio channels for three consecutive acquisitions ($P_{fa} = 10^{-4}$). (a) Channel at 89.8 MHz (SFC1). (b) Channel at 92.1 MHz (SFC2). (c) Channel at 93.8 MHz (SFC3). (d) Channel at 94.3 MHz (SFC4).



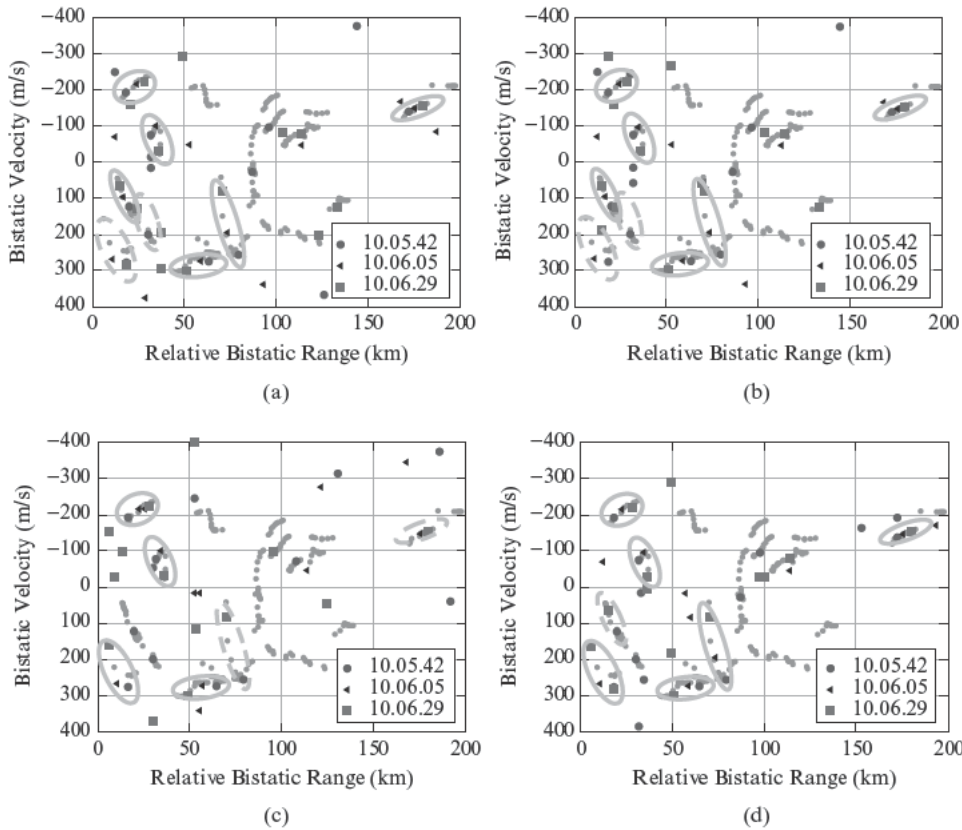


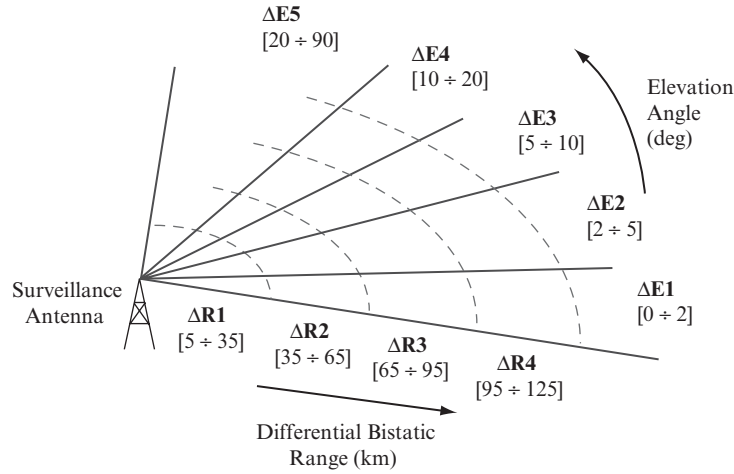
FIGURE 17-33 ■ Plot sequences in the range-velocity maps for the considered integration strategies obtained using channels at 89.8, 92.1, and 93.8 MHz ($N_{ch} = 3$) in the three subsequent acquisitions and a CA-CFAR scheme ($P_{fa} = 10^{-4}$). (a) SUM. (b) MAX. (c) MIN. (d) DEC.

Correspondingly, Figure 17-33 reports the plot sequences obtained with the different MF integration strategies after the application of a CA-CFAR detection scheme over the same sequence of three acquisitions used in Figure 17-32.

Three FM channels were integrated (89.8, 92.1, and 93.8 MHz), and the decentralized approach operates with $L = 2$. For this set of three consecutive acquisitions, the best-performing integration strategies are the SUM (i.e., the centralized linear integration), the MAX (i.e., the centralized nonlinear integration based on the maximum selection), and the DEC (i.e., the decentralized nonlinear integration) approaches that are able to detect many complete plot sequences and thus to significantly enhance the performance obtained with the SFC operation. Additional plots are detected for almost all the considered target tracks. This gives a significant advantage especially for targets that result in isolated plots or two-plot sequences when operating with a single channel, since their tracks are otherwise likely to be lost.

To provide an extensive performance comparison, the detection results obtained with the SFC operation and the MF approaches, for all the available data files, were compared with the available live ATC registrations. An extensive analysis was performed over the 300 recorded data files aiming at estimating the achieved detection probability and the actual false alarm rate. Specifically, the performance was evaluated over proper regions obtained by dividing the surveillance volume (i.e. the volume contained into the 3 dB surveillance antenna pattern) over the bistatic range and elevation dimensions, as depicted in Figure 17-34.

FIGURE 17-34 ■
Sketch of the
considered regions
for the detection
performance
evaluation.



Five MF configurations (MFC) were considered for all the considered integration strategies:

- MFC2: $N_{ch} = 2$ integrated channels (89.8, 93.8 MHz)
- MFC3: $N_{ch} = 3$ integrated channels (89.8, 93.8, 92.1 MHz)
- MFC4a: $N_{ch} = 4$ integrated channels (89.8, 93.8, 92.1, 91.3 MHz)
- MFC4b: $N_{ch} = 4$ integrated channels (89.8, 93.8, 92.1, 94.3 MHz)
- MFC5: $N_{ch} = 5$ integrated channels (89.8, 93.8, 92.1, 94.3, 91.3 MHz)

First, the false alarm rates obtained with the different decentralized and centralized integration strategies were evaluated. The performed analysis showed that the obtained values are largely comparable to the nominal desired P_{fa} value; thus, a proper control of the false alarm rate can be achieved with all the considered integration strategies [56, 57].

With the aim to analyze the target detection performance improvement of the MF operation, the decentralized integration strategy is first considered and compared with the SFC operation in Figure 17-35 for a CA-CFAR detection scheme. The figure shows the detection rates achievable with the different single/multiple-frequency approaches for three different bistatic range regions ΔR_1 , ΔR_2 , and ΔR_3 (Figure 17-34). For clarity, these values were obtained by merging the results obtained for elevation angles in the range 5–90° (ΔE_3 , ΔE_4 , and ΔE_5) and discarding the very low elevation angles where the detection performance is degraded due to the limited height of the building upon which the surveillance antenna is mounted, because it is surrounded by other structures of similar or higher height.

The decentralized strategy operates with a 2-of- N_{ch} logic for $N_{ch} < 5$ (MFC2, MFC3, MFC4a, MFC4b) and with a 3-of-5 logic in the MFC5 case. As expected, for both the SFC and the MF operation, the detection capability rapidly degrades as the relative bistatic range increases. As is apparent the decentralized integration strategy yields a detection performance improvement with respect to the SFC operation since, for any given range region, it is possible to find a MFC which is able to increase the detection rate with respect to the best performing single channel. This improvement becomes higher at farther bistatic ranges. Notice that it is quite difficult to identify the best-performing MFC since the resulting detection capability strongly depends on the single integrated channels

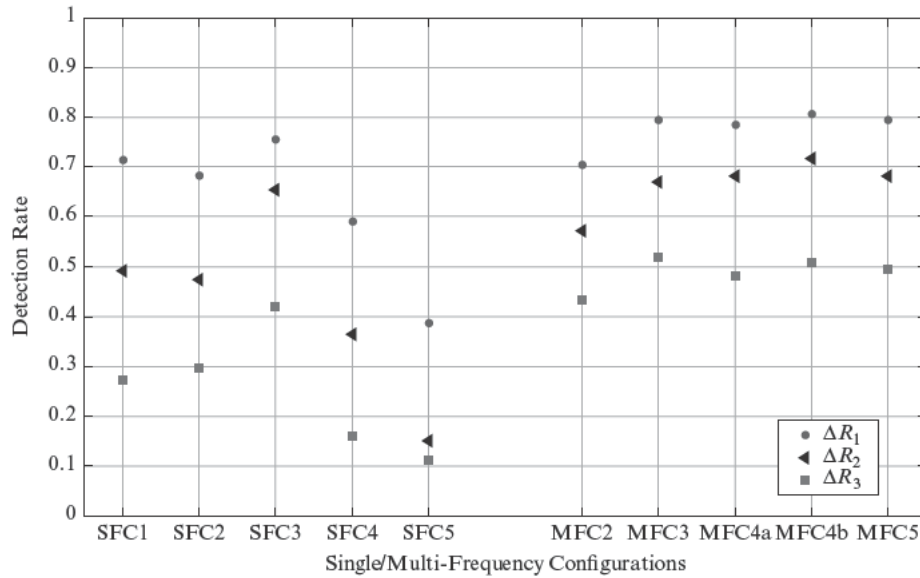


FIGURE 17-35 ■ Detection performance comparison among the different single frequency channels (SFC1-5) and the decentralized integration strategies (MFC2-5) for $P_{fa} = 10^{-3}$.

behavior, especially in a decentralized integration strategy. This is clearly apparent from a comparison of the two MF configurations based on four integrated channels (MFC4a and MFC4b). As a consequence, the increase of the number of integrated channels does not always yield a performance improvement if the additional channels show poor performance. Thus, the centralized integration strategies are analyzed and compared to exploit the SNR improvement due to the direct (linear or nonlinear) integration of the target power at the different channels.

Figure 17-36 shows the comparison of the detection performance obtained with the centralized integration strategies for a CA-CFAR integration scheme. As is apparent, the MIN approach always yields the worst performance since, for any MF configuration, the achievable detection rate generally degrades with respect to both the SUM and MAX approaches operating with the same number of integrated channels. Moreover, the best performing MF configuration with the MIN approach always yields worse performance

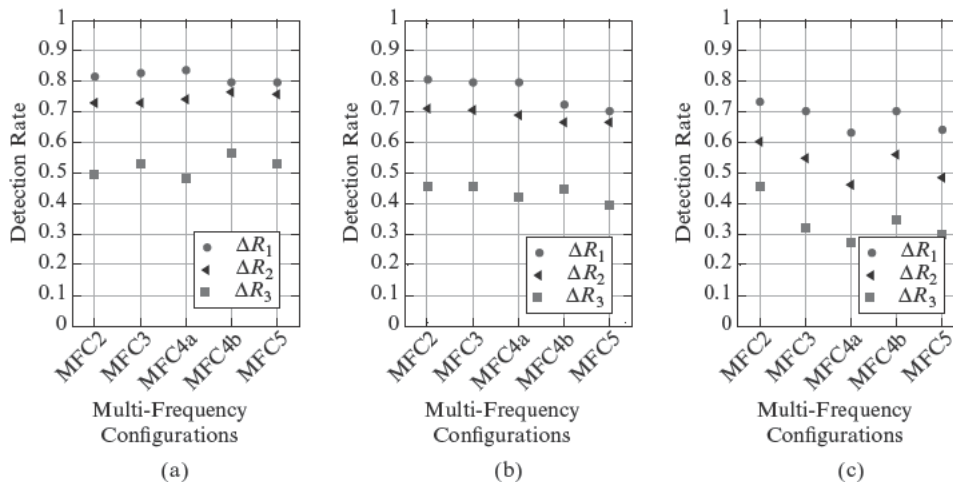
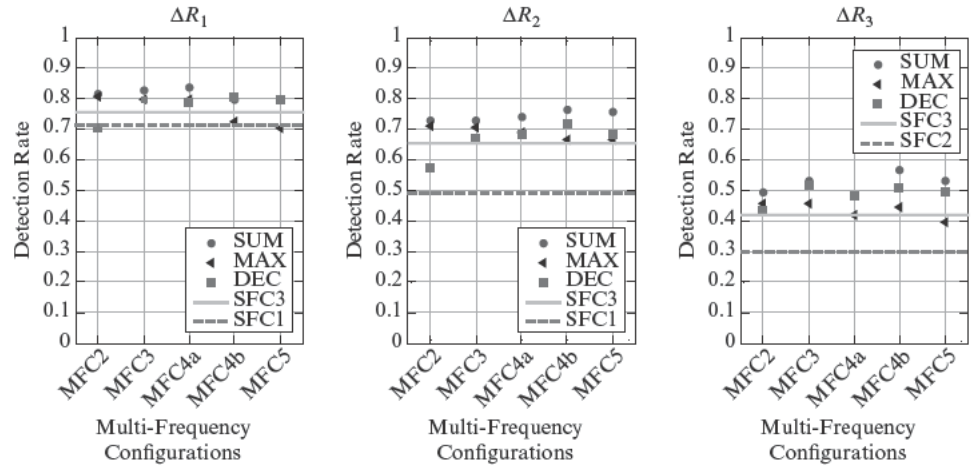


FIGURE 17-36 ■ Detection performance comparison among the centralized MF strategies ($P_{fa} = 10^{-3}$). (a) SUM. (b) MAX. (c) MIN.

FIGURE 17-37 ■
Detection
performance
comparison among
the best MF
strategies
($P_{fa} = 10^{-3}$).



than the worst MF configuration with the SUM approach. We interpret this behavior as follows: the MIN strategy is able to discard the disturbance residuals and sidelobe structures that appear at a SFC map; however, it retains the minimum SNR for any given target. This implies a large loss when operating with FM channels with largely different SNR. The SUM and the MAX integration strategies show similar performance; however, the SUM has a slight advantage at far range and when using a high number of integrated channels (i.e., four to five).

A global comparison of the proposed integration strategies is shown in Figure 17-37 for the three considered range regions. The reported results refer to the best-performing integration approaches (DEC, SUM, and MAX). In each figure the detection rates obtained with the two best-performing single channels at each range region have been added for comparison. The detection capability of the considered PBR system is upper limited due to the long baseline between the Tx and Rx. Nevertheless, the proposed MF approaches remarkably improve performance with respect to the best single channels for all the considered range regions. Unfortunately, data collected with a more favorable bistatic geometry (especially a shorter baseline) are not available; such a data set would provide both better SFC and MF performance but would not change the comparative results. The following configurations are preferred: (1) the SUM approach with any number of channels; (2) the MAX among the two or three best channels; and (3) the DEC of type 2/3 (and possibly 2/4). These configurations enhance the detection performance with respect to the best-performing single channels with all the considered MF configurations.

This performance improvement is paid in terms of computational load since the data from N_{ch} FM radio channels should be contemporaneously processed. Nevertheless, the computational burden can be limited by implementing a suboptimum technique for evaluating 2D-CCF; in fact, integrating multiple FM channels recovers the small SNR loss that occurs from suboptimum implementation of 2D-CCF with the advantage of a higher robustness of the system to time-varying characteristics of both the exploited waveform and the electromagnetic environment.

In our specific case, further increasing the number of integrated channels does not improve performance since the additional channels have been recognized to yield poorer characteristics. On the contrary, in some cases, adding those channels degrades the detection capability. Obviously, further improvement is expected by adding good-performing

channels when available. In this regard, it is potentially of great interest to explore the capability of automatically selecting the frequency channels to be integrated. Such automatic selection would allow the desired detection performance to be achieved without significantly increasing the number of channels and hence the computational load.

The MF operation was introduced as a way to counteract the time-varying detection performance of the PBR. To this purpose, noncoherent integration strategies have been adopted. However, as previously mentioned, the joint exploitation of different frequency channels might also improve the target localization capability of the PBR system.

In the first instance, a reasonable improvement in the localization accuracy can be obtained by properly combining (averaging) the measurements obtained using the signals transmitted at different carrier frequencies. This would yield a remarkable advantage, especially for the DOA estimation in low-frequency PBR when this is obtained using a simple interferometric approach [59].

Moreover a coherent combination of the 2D-CCFs obtained with the multiple received signals can be exploited aiming at increasing the bistatic range estimation accuracy. The availability of multiple channels would result in a global wider bandwidth signal to be exploited by the PBR system. However, due to the frequency separation among the different channels, this signal presents a discontinuous PSD that yields high ambiguities in the corresponding ACF. Thus, this approach cannot be used to improve the radar range resolution, namely the capability to distinguish two targets but can improve the single channel estimation of the bistatic range of a given detected target largely affected by the exploited signal ACF.

17.6.2 Exploitation of Multiple Polarimetric Passive Radar Channels

The transmitter and receiver antenna polarization play an important role in designing a PBR system. The exploited transmitter of opportunity might use vertical, horizontal, or circular polarization. Usually broadcast transmitters use linear polarization; most DAB transmissions are vertically polarized, whereas the horizontal (or mixed) polarization is also used for FM radio and DVB-T signals. Circular polarization is most often used for satellite communications. In contrast different solutions are available for the transmitter antenna used by a base station in wireless and cellular networking; in these cases, polarization diversity is usually adopted to minimize the effects of selective fading of the horizontal and vertical components of a signal. Accordingly, proper polarizations might be exploited for the reference and the surveillance antennas used by a PBR system aiming at improving its performance. Assuming the availability of at least two receiving antennas with different (orthogonal) polarizations, many approaches might be envisaged to use the conveyed diversity aiming at specific goals.

Typically, at the reference channel, the main objective is to recover a good copy of the transmitted signal; to this purpose a switching or a selection strategy might be employed to pick out the signal from the preferred antenna (e.g., the one yielding the highest DNR). Alternatively, the signals collected at the two orthogonally polarized antennas might be combined to reject the multipath contributions.

With reference to the surveillance channel, different objectives might be pursued:

1. The polarization diversity has been exploited to increase the direct signal rejection that can be obtained by selecting the antenna with cross-polarization with respect to

the transmitter antenna [7, 8]. As an alternative, it is possible to properly combine the signals received at different antennas. This solution requires the availability of multiple receiving channels, but it might be more effective when the transmitter polarization is unknown or when the direct signal polarization experiences slow changes due to the geometric and atmospheric conditions (e.g., satellite transmissions).

2. Assuming that the adopted cancellation algorithms allow a sufficient direct signal cancellation, the polarization diversity might be exploited to reject other interfering signals being transmitted in cross-polarization in the same or adjacent frequency band.
3. Target echoes show a random polarization due to the reflection of the transmitted signal on the target complex structure; thus, using a fixed polarization for the surveillance antenna might result in a significant SNR degradation depending on the target characteristics. It can therefore be expected that a proper combination of the signals received at differently polarized antennas might yield a detection performance improvement for the resulting PBR.

In the following we report some examples of the potential exploitation of different antenna polarizations to mitigate the effect of co-/adjacent-channel interferences (i.e., approach 2) in FM-based PBR [60]. To illustrate the effect of interfering signals on FM-based PBR performance, we exploit the real data set collected during the acquisition campaign carried out on April 30, 2009, at the same site and with the same geometry depicted in Figure 17-31. In this case a quad-channel experimental PBR prototype was used that allowed for synchronously sampling the analog signals from up to four input channels. Simultaneous downconversion of up to 16 arbitrary signal bands (e.g., 16 FM radio channels) is provided by four Graychip GC4016 quad digital downconverters. The described setup allows data to be collected from up to four different FM radio channels, each one from four different antennas. In the considered acquisition campaign, four linearly polarized directive antennas were used. Specifically, two of them were employed as reference antennas, one vertically polarized (Ref-V) and one horizontally polarized (Ref-H), and steered toward a transmitter located on Monte Argentario (see Figure 17-31). Similarly, the two cross-polarized surveillance antennas (Surv-V, Surv-H) were pointed at about 180° with respect to the exploited transmitter. The following results were evaluated for two FM radio channels for which the performance has been recognized to be highly affected by co-/adjacent-channel interferences (channels SFC2 and SFC4 in Table 17-8). This is shown in Table 17-9 that reports the comparison of theoretical cancellation (CA_{Theo}) and actual cancellation (CA) obtained with the cancellation algorithm described in Section 17.3. Different combinations are considered for the reference and surveillance antenna polarizations for both the considered FM channels. The reported results have been averaged over 681 consecutive acquisitions of 1.15 sec each. As is apparent, the actual cancellation values are always significantly lower than their theoretical expectations. The obtained CA loss cannot be due only to limitations of the system hardware or the exploited processing techniques. In fact, at different FM channels only small CA loss is obtained (e.g., the best-performing FM channels of Table 17-8, SFC1 and SFC3). Nevertheless, the performed analysis demonstrated that high CA loss is mainly due to the presence of interfering sources affecting the PBR performance.

The acquired signals have been demodulated and listened to before and after the cancellation stage: as expected, the reference and surveillance signals before cancellation sound identical, whereas in most cases a different radio program can be listened

TABLE 17-9 ■ Cancellation Performance against Real Data

Exploited Antennas	Frequency [MHz]	CA _{Theo} [dB]	CA [dB]	CA Loss [dB]
Ref-V/Surv-V	92.1 (SFC2)	45.4	22.2	23.2
	94.3 (SFC4)	45.8	8.5	37.3
Ref-H/Surv-V	92.1 (SFC2)	45.1	22.1	23
	94.3 (SFC4)	45.6	7.9	37.7
Ref-V/Surv-H	92.1 (SFC2)	47.5	28.3	19.2
	94.3 (SFC4)	48.1	23.6	24.5
Ref-H/Surv-H	92.1 (SFC2)	47.1	27.8	19.3
	94.3 (SFC4)	47.8	23.4	24.4

in the surveillance signal after the cancellation stage. This clearly demonstrates that the disturbance contributions in the surveillance signal have been effectively reduced by the cancellation filter. However, its performance is actually limited by the interfering signals level rather than by thermal noise only (as assumed in the evaluation of the theoretical CA). Specifically, the actual CA is limited by the direct signal-to-interference plus noise ratio at the surveillance channel ($CA \approx \text{DINR}_{\text{Surv}}$). In addition, the quality of the listened program after cancellation is better at those channels for which a higher CA loss was measured. In fact, assuming that the interference level is well above the noise floor ($\text{INR} \gg 1$), we have $CA = CA_{\text{Theo}}|_{\text{dB}} - CA_{\text{Loss}}|_{\text{dB}} \approx \text{DNR}|_{\text{dB}} - \text{INR}|_{\text{dB}}$, and we conclude that the CA loss is a measure of the interference power level. Obviously, different interfering signals may affect the desired surveillance signal. A detailed analysis of the transmitters located in the area surrounding the receiver site allowed some broadcasted radio channels to be identified that are probably responsible for interchannel interferences at the exploited FM radio channels. The identified transmissions are listed in Table 17-10 in decreasing order of expected power level at the receiver site (evaluated on the basis of the available information). They are broadcasted by FM emitters located in the surveillance area at distances largely comparable with the PBR baseline (see Figure 17-31).

TABLE 17-10 ■ Potential Interfering FM Radio Channels for the Exploited FM Radio Channels (SFC2 and SFC4)

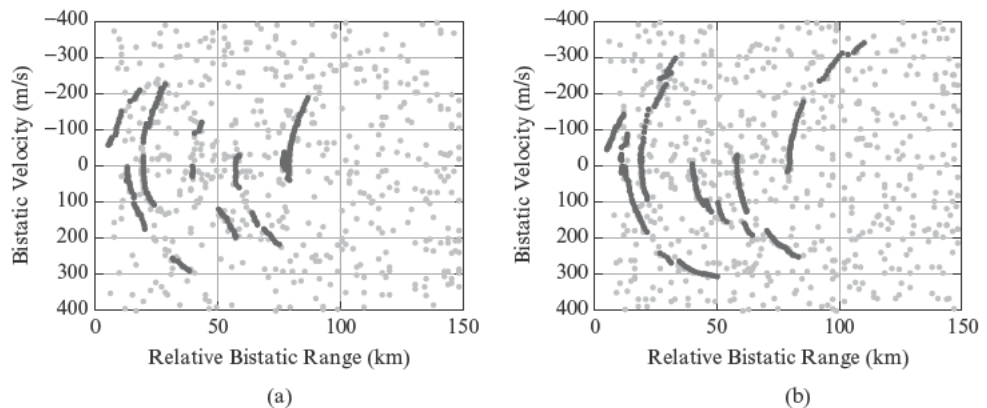
Frequency [MHz]	FM Radio Broadcast	Transmitter	Distance [km]	Polarization
92.2	RTL 102.5	TX1 (Terminillo)	105	N/A
92.0	Radio Domani	TX2 (Vignanello)	50	N/A
92.2	RTL 102.5	TX5 (Monte Gennaro)	80	V
92.0	R. 101	TX5 (Monte Gennaro)	80	V
92.0	R. Suby	TX6 (Soriano)	49	N/A
92.2	RTL 102.5	TX4 (Monte Cavo)	83	V
92.2	R. Immagine	TX7 (Velletri)	84	N/A
94.2	Subasio	TX3 (Albano)	82	V
94.5	Subasio	TX4 (Monte Cavo)	83	V
94.1	Subasio	TX1 (Terminillo)	105	H
94.4	RAI R.2	TX1 (Terminillo)	105	H

To understand the potential benefits of the exploitation of different polarizations for the receiving antennas we observe from Table 17-9 that the disturbance cancellation capability of the PBR is a function of the exploited polarizations. Specifically, changing the polarization of the reference antenna yields negligible performance variations since, as previously mentioned, the cancellation capability is limited by the DINR at the surveillance antenna. In contrast, for both the considered FM channels, using a horizontally polarized surveillance antenna allows better cancellation capability despite both the considered FM channels exploit an horizontal polarization (see Table 17-8). This clearly shows that selecting the surveillance antenna polarization to increase the direct signal rejection is not always the best choice.

The performance improvement deriving from the use of a Surv-H antenna in place of a Surv-V antenna is higher for the FM channel at 94.3 MHz (SFC4). This can be explained as follows. The interfering signals with highest power levels for this channel are transmitted in vertical polarization (see Table 17-10) so that using the opposite polarization at the receiver allows a significant reduction of the interference level, thus partially recovering the cancellation loss with respect to the theoretical value (a CA improvement higher than 15 dB is obtained). In contrast, a limited cancellation performance improvement (about 6 dB) is obtained for the FM channel at 92.1 MHz (SFC2). In this case, no information is available about the polarization exploited by the highest power interfering transmissions. It is reasonable to assume that these transmissions exploit a mixed-polarization so that the performance at the considered FM channel does not show a sharp-cutting behavior when changing the surveillance antenna polarization.

As a consequence, it is expected that, for the FM radio channel at 94.3 MHz (SFC4), using a horizontal polarization yields a reduced system noise floor and, consequently, a better detection capability. This is clearly shown in Figure 17-38 that reports the detection results obtained for 150 consecutive acquisitions on the same bistatic range-velocity map after the application of a CA-CFAR detection scheme with a nominal false alarm probability $P_{fa} = 10^{-4}$, when different polarizations are exploited for the surveillance antenna. Specifically, Figure 17-38a is obtained using the data collected at the Surv-V antenna, whereas Figure 17-38b refers to the Surv-H antenna. The light plots represents the raw detections, whereas dark plots are used for the target tracks identified by using a conventional tracking algorithm over the bistatic range-velocity plane. According to the results obtained in terms of cancellation capability, significantly better detection performance can be achieved for the FM channel at 94.3 MHz when the Surv-H antenna is employed. In fact, additional

FIGURE 17-38 ■
Plot sequences in the range-velocity map for 150 consecutive acquisitions for FM channel at 94.3 MHz (SFC4) with (a) Ref-H/Surv-V and (b) Ref-H/Surv-H.



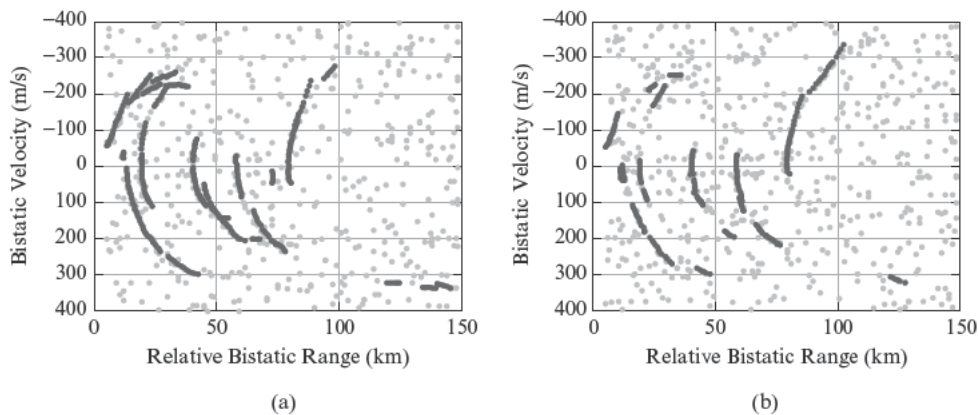


FIGURE 17-39 ■ Plot sequences in the range-velocity map for 150 consecutive acquisitions for FM channel at 92.1 MHz (SFC2) with (a) Ref-H/Surv-V and (b) Ref-H/Surv-H.

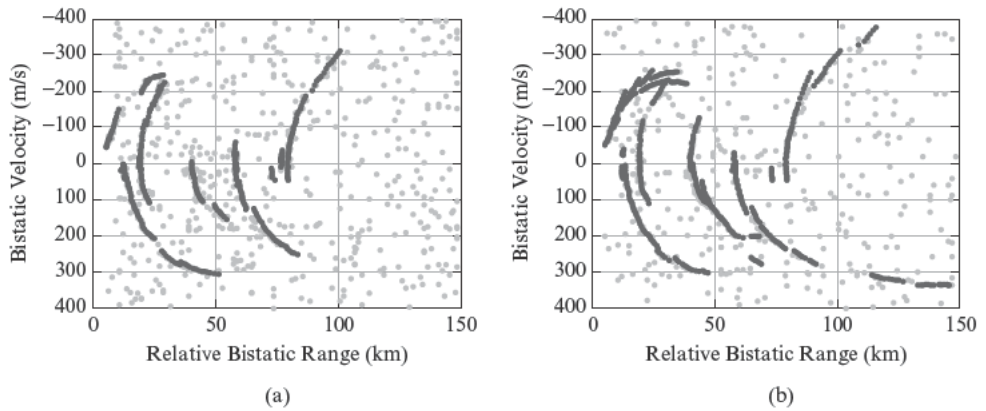
target tracks can be identified with respect to the Surv-V case. Furthermore, for a given target track, a higher number of plots is obtained thus making the tracking stage easier.

Similar results are shown in Figures 17-39a–b for the FM channel at 92.1 MHz (SFC2) with Surv-V and Surv-H, respectively, for the same set of acquisitions. In this case, comparable detection performances are achieved using different surveillance antenna polarizations. However, regardless of the exploited polarization, the FM channel at 92.1 MHz yields better performance with respect to the FM channel at 94.3 MHz since it is less affected by interferences (the lower interference level is demonstrated by the lower CA loss in Table 17-9). Also, the exploitation of different polarizations allows different target plots within the same target track to be detected. Therefore, in this case further performance improvement is expected if proper techniques are devised to merge the diversity of information conveyed by different polarizations.

A preliminary approach can be obtained by considering a linear integration operating by incoherently summing with equal weights the range-velocity maps (after square root detector) obtained using different polarizations. The CFAR detection threshold to be applied should be properly modified to control the resulting P_{fa} [56, 57]. The expected benefits of such approach might be explained as follows. By changing the surveillance antenna polarization, the desired signal might be differently affected by interferences since the interfering signals exploiting different polarizations are differently weighted. Assuming that the interfering signals yield a different power distribution over the range-Doppler plane, the linear integration would yield a reduced background level over the integrated map. In addition, assuming that the target returns are randomly polarized, the proposed approach would enhance on average the SNR, with respect to the use of a single fixed polarization, thus increasing the detection capability of the PBR system.

This is confirmed in Figures 17-40a–b. They show the detection performance obtained after polarimetric integration for the FM channels at 94.3 MHz (SFC4) and 92.1 MHz (SFC2), respectively. By comparing these results with those reported in Figure 17-38 and Figure 17-39 for same set of acquisitions, it is apparent that the number and the completeness of the detected target tracks have increased. As expected, the performance improvement at 94.3 MHz is limited with respect to the best-performing surveillance antenna polarization (Surv-H); however, a dramatic improvement is obtained with respect to the Surv-V case representing a clear advantage of the proposed technique when no a priori information is available about the interference scenario or when it exhibits a time-varying behavior. In contrast, for an FM channel not yielding a sharp-cutting

FIGURE 17-40 ■ Plot sequences in the range-velocity map for 150 consecutive acquisitions after integration of Surv-H and Surv-V. (a) FM channel 94.3 MHz (SFC4). (b) FM channel 92.1 MHz (SFC2).



polarimetric behavior (e.g., 92.1 MHz), the proposed polarimetric integration allows a significant improvement in terms of detection capability. For example, in Figure 17-40b, a quite complete plot sequence is detected at ranges higher than 120 km, while it was only partially identified with the single polarization operation.

17.6.3 Adaptive Antenna Array for Passive Radar

A PBR might exploit directive antennas, characterized by a high front-to-back ratio, to obtain a partial rejection of the direct signal from the transmitter. However, directive antennas provide only a limited angular coverage for the radar surveillance and do not cover a very wide air space region. This drawback can be avoided by using a phased array composed of a set of omnidirectional elements. As long as the elements are properly spaced, a directive pattern may be synthesized by jointly processing the signals received at each element. This also allows the beam to be electronically steered in all directions or, better, a set of directional beams to be generated that globally cover the whole air space of interest. Furthermore, the considered array may be provided with spatial adaptivity, so that the clutter/multipath cancellation filter can benefit from space-time processing. Obviously this would require the availability of multiple coherent receiving channels. However, in this case the same array might be used to collect the transmitted signal by synthesizing a beam pointed toward the transmitter of opportunity. Even for the reference signal recovery, adaptive techniques can be used to protect the system against multipath contributions.

Aiming at using systems with limited complexity, in the following we describe effective criteria to identify appropriate configuration for the array, able to provide good performance using a limited number of elements [61]. To this purpose we refer to some significant pattern parameters, such as the 3 dB aperture and the PSLR. Consider a planar array of M elements, and let (x_i, y_i) and $g_i(\theta)$ be the position and the azimuth pattern of the i -th element, respectively, being θ the angle formed with the positive y-axis. Then, the target steering vector $\mathbf{s}_\theta(\theta)$ can be written as $\mathbf{s}_\theta(\theta) = [s_\theta^{(1)}, s_\theta^{(2)}, \dots, s_\theta^{(M)}]^T$, where

$$s_\theta^{(i)}(\theta) = g_i(\theta) \exp \left\{ j \frac{2\pi}{\lambda} (x_i \sin \theta + y_i \cos \theta) \right\} \quad i = 1, \dots, M \quad (17.40)$$

The DOA θ_T of a monochromatic signal with unknown complex amplitude a may be estimated by maximizing the generalized likelihood (GL). Assuming that the disturbance

can be modeled as additive white Gaussian noise with variance σ_n^2 , the resulting DOA estimation accuracy is given by

$$\begin{aligned}\sigma_{\theta_T}(\theta_T) &= \sqrt{E\left\{|\hat{\theta}_T - \theta_T|^2\right\} - |E\{\hat{\theta}_T - \theta_T\}|^2} \\ &= \left\{ 2 \frac{|a|^2}{\sigma_n^2} \left[\|\dot{\mathbf{s}}_{\theta}(\theta_T)\|^2 - \frac{|\dot{\mathbf{s}}_{\theta}^H(\theta_T) \mathbf{s}_{\theta}(\theta_T)|^2}{\|\mathbf{s}_{\theta}(\theta_T)\|^2} \right] \right\}^{-1/2}\end{aligned}\quad (17.41)$$

Using omnidirectional antennas, that is, $g_i(\theta) = g$ ($i = 1, \dots, M$), the estimation accuracy in (17.41) becomes

$$\sigma_{\theta_T}(\theta_T) = \left[2 \frac{|a|^2}{\sigma_n^2} g^2 \frac{4\pi^2}{\lambda^2} f(\theta_T) \right]^{-1/2} \quad (17.42)$$

where the dependence on the DOA θ_T is embedded in the term $f(\theta_T)$ given by

$$f(\theta_T) = \sum_{i=1}^M (x_i \cos \theta_T - y_i \sin \theta_T)^2 - \frac{1}{M} \left[\sum_{i=1}^M (x_i \cos \theta_T - y_i \sin \theta_T) \right]^2 \quad (17.43)$$

If the following symmetry conditions hold for the planar array

$$\sum_{i=1}^M x_i = 0 \quad \& \quad \sum_{i=1}^M y_i = 0 \quad \& \quad \sum_{i=1}^M x_i^2 = \sum_{i=1}^M y_i^2 \quad \& \quad \sum_{i=1}^M x_i y_i = 0 \quad (17.44)$$

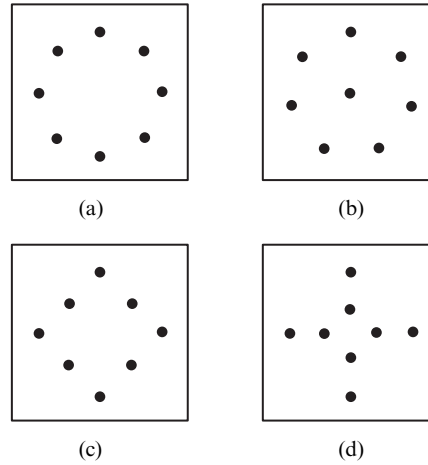
equation (17.43) becomes

$$f(\theta_T) = \frac{1}{2} \sum_{i=1}^M (x_i^2 + y_i^2) = \sum_{i=1}^M x_i^2 = \sum_{i=1}^M y_i^2 = \frac{1}{2} \sum_{i=1}^M R_i^2 \quad (17.45)$$

where R_i is the Euclidean distance of the i -th element from the origin of the reference system. In this case, the estimation accuracy in (17.42) is independent of the DOA θ_T . This is a desirable characteristic when the considered application does not require specific directions to be preferred. Thus, in the following we consider only symmetric array configurations, namely, array configurations for which the symmetry conditions in (17.44) hold.

The selection of the most appropriate configuration for passive radar applications involves the choice of the number M of antenna elements, the identification of its shape and the determination of the array radius R , defined as the radius of the smallest circle containing all the array elements. Considering an array with a relatively small number of elements, a suitable configuration might be selected on the basis of its pattern characteristics, such as the 3 dB aperture and the PSLR. In fact, a selection based only on the estimation accuracy in (17.41) would not be appropriate since such expression is significant only in the absence of ambiguities, namely, when grating lobes do not occur. A parameter related to the presence of grating lobes, such as the PSLR, has therefore to be considered together with the estimation accuracy. Moreover, since this latter parameter is

FIGURE 17-41 ■
Symmetric array
configurations for
 $M = 8$.



directly related to the pattern main lobe width, the 3 dB aperture may be considered in place of the accuracy.

An extensive analysis of many different configurations showed that it is not possible to synthesize array patterns characterized by both a narrow beamwidth and a satisfactory PSLR by using less than $M = 8$ elements. Some reasonable symmetric array configurations of $M = 8$ elements are shown in Figure 17-41 (see also [62]); correspondingly, Figures 17-42a–b show the comparison of the 3 dB aperture and the PSLR, respectively, as a function of the ratio R/λ for the different array configurations of Figure 17-41. As both parameters vary with the pointing direction θ_0 , the worst case is reported. A joint analysis of the two plots allows to directly select an appropriate configuration. In particular, the pattern synthesized by the configuration in Figure 17-41b is characterized by a reasonable PSLR (13 dB), also when the 3 dB aperture is less than 50° . This characteristic makes this configuration especially attractive, and the radius R of the array is selected to be 0.44λ . For example, for FM radio-based PBR, the wavelength of the signal is comprised between $\lambda_{\min} = 2.78$ m (108 MHz) and $\lambda_{\max} = 3.41$ m (88 MHz); thus, the radius of the array has to be chosen as a compromise value, with special attention to λ_{\min} .

It is well-known that it is possible to synthesize an array pattern with a higher PSLR by properly weighting the single array elements. However, in the general case the application of a taper function intrinsically destroys the uniform angular performance. Special considerations apply to the array configuration in Figure 17-41b, since in this case there is a simple way to apply something similar to a taper function by weighting differently the array central element and the circle of seven elements.

The performed analysis demonstrated that a higher PSLR can be obtained by properly tapering only the array central element. It also allowed us to identify the tapering factor α that maximizes the PSLR. Figure 17-43 shows the pattern synthesized with the array configuration in Figure 17-41b and the described radial taper. As is apparent, a PSLR greater than 20 dB is obtained with only a slight broadening of the antenna pattern main lobe. This tapered configuration is especially effective for PBR purposes, since it improves the direct signal rejection up to 7 dB, when the main beam is not steered 180° away from the transmitter of opportunity.

As the received power of the direct signal is several dBs, or even tenths of dBs, greater than the power received from a target echo, even using amplitude tapers, the direct

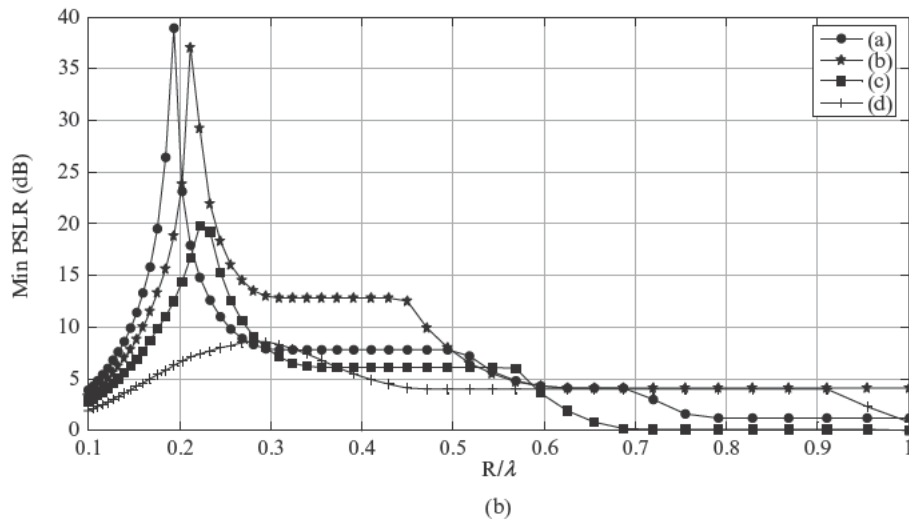
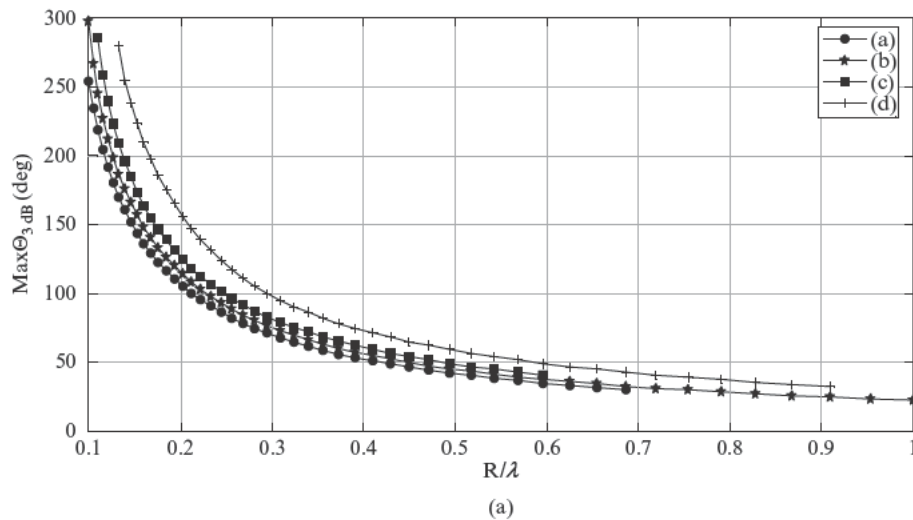
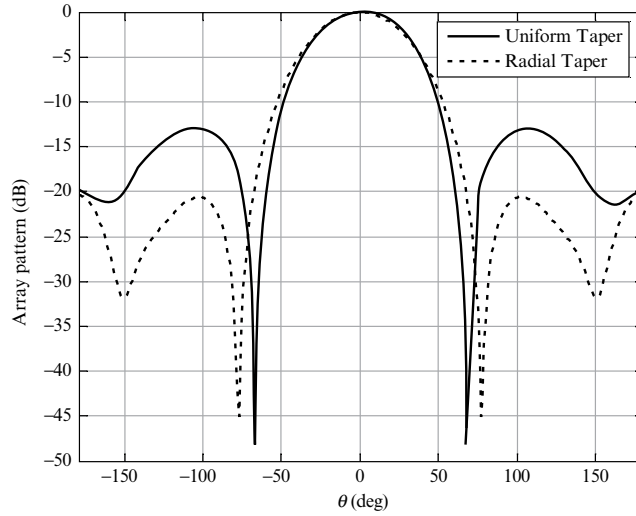


FIGURE 17-42 ■
Pattern
characteristics for
the array
configurations
depicted in
Figure 17-41.
(a) 3 dB aperture.
(b) PSLR.

signal attenuation might not be high enough, and an adaptive temporal cancellation is typically used to remove its effect together with the clutter to provide the desired detection performance (see Section 17.3). However, when the direct signal is especially strong (e.g., when the passive radar receiver is very close to the transmitter of opportunity), temporal cancellation techniques following the antenna sidelobe attenuation of the direct signal might still be unsatisfactory. In this case spatially adaptive techniques may be considered to reject the disturbance, thus yielding increased detection performance. A clear demonstration of potentiality of the spatial adaptivity is reported also in [11], for a specific passive radar application in the HF bandwidth.

Antenna-based adaptive cancellation techniques provide a rejection of the disturbance by means of adaptive beamforming. An estimate of the disturbance covariance matrix \mathbf{M} is required to properly weight the received signals. Three approaches to adaptive cancellation are considered and discussed in the following.

FIGURE 17-43 ■
Array patterns for
array configuration
in Figure 17-41b.



Optimum Filter Assuming that the disturbance covariance matrix \mathbf{M} is known, the optimum weights vector that maximizes the signal-to-disturbance ratio is given by [63]

$$\mathbf{w}^H(\theta_0) = \mathbf{s}_\theta^H(\theta_0) \mathbf{M}^{-1} \quad (17.46)$$

where θ_0 is the direction toward which the beam is pointed. In practical applications, the covariance matrix \mathbf{M} is replaced with its estimated version obtained from the received data. This is quite appropriate in the PBR case given that the level of both the direct signal and its multipath reflections is well above that of the useful signal. After combining the received signals according to (17.46), an adaptive temporal cancellation might be performed only to cancel disturbance residuals not adequately suppressed by the spatial adaptive filter. Then the 2D-CCF is evaluated, and a proper CFAR threshold is applied to obtain a detection map. When applying the optimum filter in the presence of a strong multipath reflection from a given direction, a null is imposed in the synthesized adaptive pattern at this direction. This also cancels all target echoes coming from the same DOA.

Principal Eigenvalue Approach A way to circumvent the weakness of the optimum filter for targets at the same DOA of strong multipaths is to reduce the number of DOF available to the spatial adaptivity so that only a single null can be imposed by the adaptive spatial filter. Since the direct signal is by far the strongest component of the overall disturbance, it is obvious that its effect dominates the spatial covariance matrix and specifically determines its principal eigenvector \mathbf{q}_{\max} . Therefore, by estimating only the principal eigenvector of the spatial disturbance covariance matrix, it is possible to strongly attenuate the direct signal by projecting the target steering vector $\mathbf{s}_\theta(\theta)$ on the subspace orthogonal to such eigenvector. The resulting weights vector is given by

$$\mathbf{w}_{PE}^H(\theta_0) = \mathbf{s}_\theta^H(\theta_0) (\mathbf{I} - \mathbf{q}_{\max} \mathbf{q}_{\max}^H) \quad (17.47)$$

Obviously, the disturbance subspace can be extended by including the eigenvectors corresponding to the highest eigenvalues. This technique leaves some multipath reflections

uncanceled, but their removal can be obtained with the adaptive temporal filter. A reference beam, orthogonal to the surveillance one, has to be digitally synthesized and yields the reference signal to be used by the adaptive temporal filter. Finally, the 2D-CCF is evaluated and the CFAR detection map is built.

Sidelobe Canceller An alternative approach can be obtained using the sidelobe canceller concept [63]. For instance, with reference to the specific array configuration previously considered (see Figure 17-41b), we might exploit seven of the eight elements of the array to form the quiescent beam and the central element as an omnidirectional retrodirective beam. Thus, the resulting weights vector for the sidelobe canceller is given by

$$\mathbf{w}_{SLC}^H(\theta_0) = \mathbf{s}_{\theta}^H(\theta_0) \mathbf{T}^H (\mathbf{T} \mathbf{T}^H)^{-1} \mathbf{T} \quad (17.48)$$

where the matrix \mathbf{T} is defined as

$$\mathbf{T} = \begin{bmatrix} s_{\theta}^{*(1)}(\theta_0) & s_{\theta}^{*(2)}(\theta_0) & \cdots & s_{\theta}^{*(7)}(\theta_0) & 0 \\ 0 & 0 & \cdots & 0 & s_{\theta}^{*(8)}(\theta_0) \end{bmatrix} \quad (17.49)$$

where $s_{\theta}^{(i)}(\theta_0)$ is the i -th element of the target steering vector ($i = 8$ refers to the central element). Notice that in this case a 2×2 matrix in place of a $M \times M$ matrix has to be inverted. Even in this case, the array pattern strongly attenuates only the signals coming from DOA close to the direct signal; therefore, the multipath reflections have to be canceled by the adaptive temporal filter.

To show the effectiveness of the considered spatial cancellation techniques, a performance comparison is presented below for a simulated FM-based PBR scenario exploiting the array configuration in Figure 17-41b. The results have been averaged over several runs. Aiming at assessing the detection performance of the considered techniques, PBR scenarios with a relatively large number of targets have been considered. A total of 16 targets have been injected in the simulated data with target/receiver distance ranging from 10 to 100 km, DOA between 150° and 180° , bistatic velocities in the range $[-150, 200]$ m/s, and SNR between -35 and -55 dB. As expected, in the absence of multipath/clutter contributions, the three adaptive approaches yield comparable performance allowing detection of 13 of 16 targets. However, when such contributions are considered in the simulated data, the optimum filter experiences some detection loss. Specifically, it fails to detect targets with DOA close to the DOA of the multipath returns. In this case, the principal eigenvalue and the sidelobe canceller approach yield more robust detection performance. Figure 17-44 reports the detection results obtained for a simulated scenario containing several stationary scatterers with clutter-to-noise ratio between 25 and 30 dB.

As is apparent, the optimum filter (Figure 17-44a) yields an effective removal of all the undesired disturbance contributions that also involves many of the injected targets. In contrast, using a reduced number of DOF, as required by the principal eigenvalue approach (Figure 17-44b) and the sidelobe canceller (Figure 17-44c), prevents undesired target cancellation at the DOA of the strong multipath echoes, thus yielding remarkable detection performance.

The many possibilities opened by exploiting adaptive array of antennas for PBR signal processing offer a wide range of possibilities for both detection (where the previous example results have been shown) and for DOA estimation, where the antenna array might allow for increased accuracy and robustness.

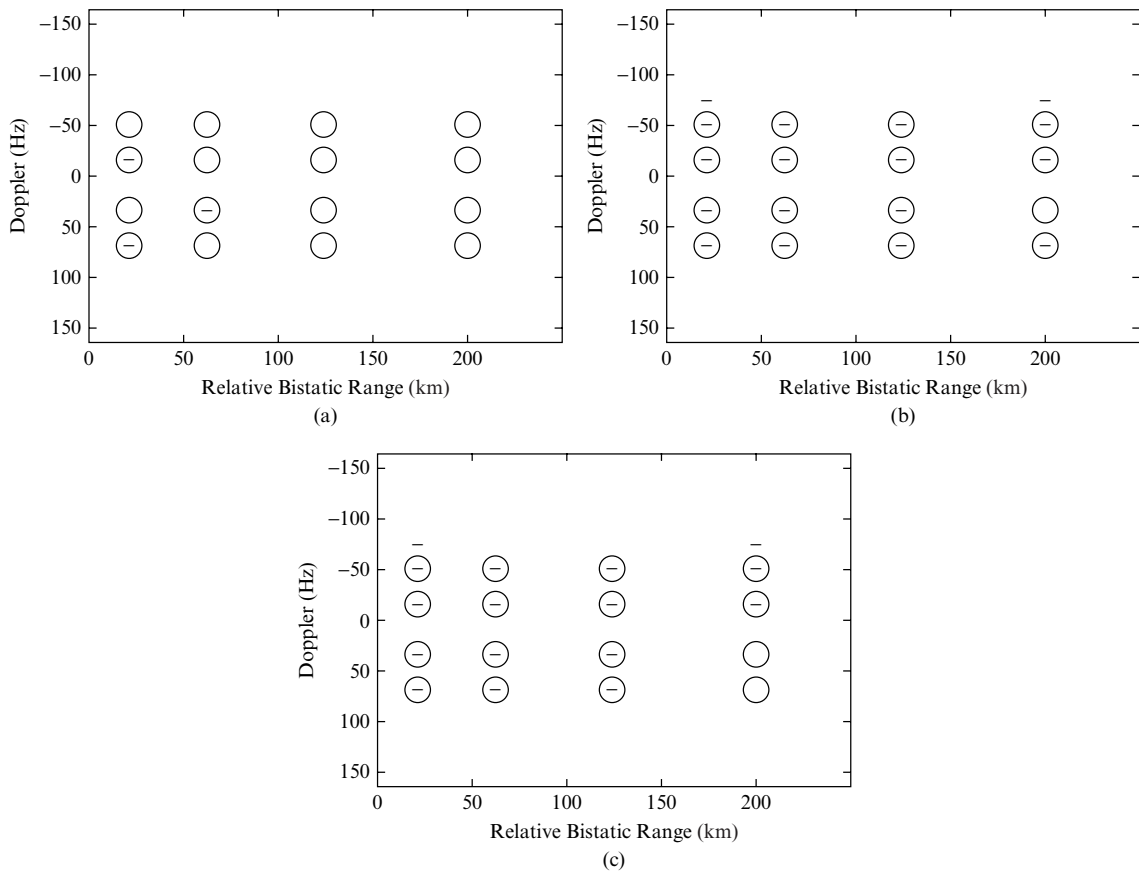


FIGURE 17-44 ■ Detection results for a simulated scenario including the direct signal and several multipath contributions for (a) optimum filter, (b) principal eigenvalue approach, (c) sidelobe canceller.

17.7 SUMMARY

In this chapter, after introducing the main scheme for a passive bistatic radar signal processor, advanced processing techniques have been illustrated to perform optimized detection with such sensors. In particular, the most essential steps of 2D-CCF evaluation and interference cancellation have been described in detail, showing their optimization in terms of computation resources and effectiveness. Thereafter, many specific problems related to the PBR operation were considered by providing advanced processing schemes for them and showing the results achievable against real data sets.

In particular, we have addressed the problems of cleaning the reference signal when exploiting sources of opportunity related to both analog and digital transmission; the problem of controlling the sidelobes of the 2D-CCF to avoid weak target masking has been addressed with reference to short-, medium-, and long-range surveillance, referring, respectively, to DSSS WiFi waveforms, OFDM WiMax waveforms, and DVB-T waveforms. We have merely brushed the surface of the potential for PBR operation when using multichannel receiving systems, with specific reference to PBR sensors characterized by multiple-frequency, polarimetric, or adaptive antenna array capability.

While far from being comprehensive, some of the already explored potentialities of advanced signal processing have been demonstrated with reference to different types of PBR. Many further improvements are expected to be developed in the next few years to take advantage of the full potential of passive radar sensors.

17.8 | ACKNOWLEDGMENTS

We are glad to acknowledge that some parts of this chapter were investigated while working under support from Finmeccanica, by means of SESM and SELEX-SI, and from the European Union – 7th Framework Program inside the two Projects: ARGUS 3D (AiR GUIdance and Surveillance 3D) – Theme Security, and ATOM (Airport detection and Tracking Of dangerous Materials by passive and active sensors arrays) – Theme Transport. We also gratefully acknowledge the collaboration of Dr. C. Bongioanni, Dr. R. Cardinali, P. Falcone, D. Langellotti, and M. Villano for parts of the reported results.

17.9 | FURTHER READING

This chapter is specifically devoted to the advanced signal processing techniques that are applied at the PBR receiver. For a more detailed global system view of the PBR the reader is referred to the following:

- Willis, N., and Griffiths, H., *Advances in Bistatic Radar*, SciTech Publishing Inc., Raleigh, NC, 2007.

where these aspects are fully covered together with the achievable performance.

A wide spectrum of topics being researched in the area of passive radar systems can be found in the following:

- Special Issue on Passive Radar Systems – IEE Proceedings on Radar, Sonar and Navigation, June 2005, Vol. 152, Issue 3, pp. 106–223.

This Special Issue, Edited by Dr. Paul Howland, includes papers addressing the use of a variety of transmitters of opportunity – FM radio, digital audio broadcast (DAB), digital video broadcast (DVB), global navigation satellite systems and cell-phone base-stations. Other papers examine the main signal processing challenges and look at the issues of deployment modelling and performance prediction. Finally, a number of papers address the problems of optimally siting the radar receivers for passive radar and bearings-only systems.

For additional details on the techniques presented in this chapter, the reader is referred to the papers by the authors as duly referenced in the text.

17.10 | REFERENCES

- [1] Special no. on Passive Radar Systems, *IEE Proceedings on Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 106–223, June 2005.
- [2] Griffiths, H.D. and Baker, C.J., “Passive Coherent Location Radar Systems. Part 1: Performance Prediction,” *IEE Proceedings on Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 153–159, June 2005.

- [3] Baker, C.J., Griffiths, H.D., and Papoutsis, I., "Passive Coherent Location Radar Systems. Part 2: Waveform Properties," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 160–168, June 2005.
- [4] Howland, P.E., Maksimiuk, D. and Reitsma, G., "FM Radio Based Bistatic Radar," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 152, no. 3, June 2005, pp. 107–115.
- [5] Bongioanni, C., Colone, F., and Lombardo, P., "Performance Analysis of a Multi-Frequency FM Based Passive Bistatic Radar," in *IEEE Radar Conference*, Rome, Italy, May 26–30, 2008.
- [6] Poullin, D., "Passive Detection Using Broadcasters (DAB, DVB) with CODFM modulation," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 143–152, June 2005.
- [7] Saini, R., Cherniakov, M., and Lenive, V., "Direct Path Interference Suppression in Bistatic System: DTV Based Radar," in *Proceedings of the International Radar Conference*, pp. 309–314, 2003.
- [8] Coleman, C. and Yardley, H., "Passive Bistatic Radar Based on Target Illuminations by Digital Audio Broadcasting," *IET Radar, Sonar & Navigation*, vol. 2, no. 5, pp. 366–375, 2008.
- [9] Saini, R. and Cherniakov, M., "DTV Signal Ambiguity Function Analysis for Radar Application," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 133–142, June 2005.
- [10] Bongioanni, C., Colone, F., Langellotti, D., Lombardo, P., and Bucciarelli, T., "A New Approach for DVB-T Cross-Ambiguity Function Evaluation," in *Proceedings of European Radar Conference*, Rome, Italy, September 30–October 2, 2009.
- [11] Fabrizio, G., Colone, F., Lombardo, P., and Farina, A., "Adaptive Beamforming for High-Frequency Over-the-Horizon Passive Radar," *IET Radar Sonar and Navigation*, vol. 3, no. 4, pp. 384–405, August 2009.
- [12] Thomas, J.M., Griffiths, H.D., and Baker, C.J., "Ambiguity Function Analysis of Digital Radio Mondiale Signals for HF Passive Bistatic Radar," *Electronic Letters*, vol. 42, No. 25, pp. 1482–1483, December 2006.
- [13] Tan, D., Sun, H, Lu, Y., Lesturgie, M., and Chan, H., "Passive Radar Using Global System for Mobile Communication Signal: Theory, Implementation and Measurements," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 116–123, June 2005.
- [14] Zemhari, R., Nickel, U., and Wirth, W.D., "GSM Passive Radar for Medium Range Surveillance," in *Proceedings of European Radar Conference*, Rome, Italy, September 30–October 2, 2009.
- [15] Colone, F., Falcone, P., and Lombardo, P., "Ambiguity Function Analysis of WiMAX Transmissions for Passive Radar," *IEEE Int. Radar Conference*, Washington, DC, May 10–14, 2010.
- [16] Colone, F., Woodbridge, K., Guo, H., Mason, D., and Baker, C.J., "Ambiguity Function Analysis of Wireless LAN Transmissions for Passive Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 240–264, January 2011.
- [17] Lombardo, P., Colone, F., and Cardinali, R., "Detection and Identification of Dangerous Materials for Airport Security," *Homeland Security Facets: Threats, Countermeasures, and the Privacy No.*, Ed. G. Franceschetti and M. Grossi, Artech House, Boston, MA, 2011.
- [18] Colone, F., Falcone, P., Bongioanni, C., and Lombardo, P., "WiFi-Based Passive Bistatic Radar: Data Processing Schemes and Experimental Results", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1061–1079, April 2012.
- [19] Guo, H., Coetzee, S., Mason, D., Woodbridge, K., and Baker, C., "Passive Radar Detection Using Wireless Networks," in *Proceedings of the Int. Conference on Radar Systems*, Edinburgh, UK, pp. 1–4, 2007.

- [20] Chetty, K., Smith, G., Guo, H., and Woodbridge, K., "Target Detection in High Clutter Using Passive Bistatic WiFi Radar," in *Proceedings of IEEE Radar Conference*, pp. 1–5, May 4–8, 2009.
- [21] Falcone, P., Colone, F., Bongioanni, C., and Lombardo, P., "Experimental Results for OFDM WiFi-Based Passive Bistatic Radar," *IEEE Int. Radar Conference*, Washington, DC, May 10–14, 2010.
- [22] Cherniakov, M., Zeng, T., and Plakidis, E., "Galileo Signal-Based Bistatic System for Avalanche Prediction," in *Proceedings IGARSS03*, Toulouse, France, pp. 784–786, July 2003.
- [23] Cherniakov, M., Saini, R., Zuo, R., and Antoniou, M., "Space-Surface Bistatic Synthetic Aperture Radar with Global Navigation Satellite System Transmitter of Opportunity—Experimental Results," *IET Proceedings on Radar, Sonar & Navigation*, vol. 1, no. 6, pp. 447–458, November 2007.
- [24] Cherniakov, M., Nezlin, D., and Kubin, K., "Air Target Detection via Bistatic Radar Based on LEOS Communications System," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 149, no. 1, pp. 33–38, 2002.
- [25] Cristallini, D., Caruso, M., Falcone, P., Langellotti, D., Bongioanni, C., Colone, F., Scafè, S., and Lombardo, P., "Space-Based Passive Radar Enabled by the New Generation of Geostationary Broadcast Satellites," *2010 IEEE Aerospace Conference*, Big Sky, Montana, March 6–13, 2010.
- [26] Gill, L.P., Grenier, D., Chouinard, J.Y., "Use of XMTM Radio Satellite Signal as a Source of Opportunity for Passive Coherent Location," *IET Radar Sonar Navigation*, vol. 5, no. 5, pp. 536–544, 2011.
- [27] Morabito, A.N., Meyer, M.G., and Sahr, J.D., "Improved Computational Performance for Distributed Passive Radar Processing through Channelised Data," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 179–184, June 2005.
- [28] Langellotti, D., Colone, F., Bongioanni, C., and Lombardo, P., "Comparative Study of Ambiguity Function Evaluation Algorithms for Passive Radar," *International Radar Symposium*, Hamburg, Germany, September 9–11, 2009.
- [29] Colone, F., O'Hagan, D.W., Lombardo, P., and Baker, C.J., "A Multistage Processing Algorithm for Disturbance Removal and Target Detection in Passive Bistatic Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 2, pp. 698–721, April 2009.
- [30] Kulpa, K.S. and Czekaa, Z., "Masking Effect and Its Removal in PCL Radar," *IEE Proceedings Radar Sonar Navigation*, vol. 152, no. 3, pp. 174–178, June 2005.
- [31] Axelsson, S.R.J., "Improved Clutter Suppression in Random Noise Radar," in *URSI 2005 Commission F Symposium on Microwave Remote Sensing of the Earth, Oceans, Ice, and Atmosphere*, April 2005.
- [32] Gunner, A., Temple, M.A., and Claypoole, Jr., R.J., "Direct-Path Filtering of DAB Waveform from PCL Receiver Target Channel," *Electronic Letters*, vol. 39, no. 1, pp. 1005–1007, 2003.
- [33] Cardinali, R., Colone, F., Ferretti, C., and Lombardo, P., "Comparison of Clutter and Multipath Cancellation Techniques for Passive Radar," in *IEEE Radar Conference*, Boston, MA, March 2007.
- [34] Colone, F., Cardinali, R., and Lombardo, P., "Cancellation of Clutter and Multipath in Passive Radar Using a Sequential Approach," in *IEEE 2006 Radar Conference*, Verona, NY, pp. 393–399, April 24–27, 2006.
- [35] Haykin, S., *Adaptive Filter Theory*, 4th ed., Prentice Hall, Upper Saddle River, NJ, 2002.

- [36] Lombardo, P., Colone, F., Bongioanni, C., Lauri, A., and Bucciarelli, T., "PBR Activity at INFOCOM: Adaptive Processing Techniques and Experimental Results," *IEEE Radar Conference*, Rome, Italy, May 26–30, 2008.
- [37] Proakis, J.G. and Salehi, M., *Digital Communications*, 5th ed., McGraw Hill, New York, 2007.
- [38] Godard, D.N., "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Transactions on Communications*, vol. 28, 1867–1875, 1980.
- [39] Treichler, J.R. and Agee, B.G., "A New Approach to Multipath correction of Constant Modulus Signal," *IEEE Transaction on Acoustics, Speech, and Signal Processing*, vol. 31, no. 2, pp. 459–472, April 1983.
- [40] Zhou, C., Sahr, J.D., Meyer, M.G., Gidner, D.M., "Ground Clutter Subtraction Algorithm for VHF Passive Radar Observation of the Upper Atmosphere," *URSI*, Maastricht, August 2002.
- [41] Colone, F., Cardinali, R., Lombardo, P., Crognale, O., Cosmi, A., Lauri, A., and Bucciarelli, T., "Space-Time CMA for Multipath Removal on the Reference Signal Exploited by Passive Bistatic Radar," *IET Radar Sonar and Navigation*, vol. 3, no. 3, pp. 253–264, June 2009.
- [42] Cardinali, R., Colone, F., Lombardo, P., Crognale, O., Cosmi, A., and Lauri, A., "Mutipath Cancellation on Reference Antenna for Passive Radar which Exploits FM Transmission," *IET International Radar Conference*, Edinburgh, UK, October 15–18, 2007.
- [43] Shynk, J.J. and Gooch, R.P., "The Constant Modulus Array for Cochanel Signal Copy and Direction Finding," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 652–660, March 1996.
- [44] Sharma, S.K. and Shafai, L., "Beam Focusing Properties of Circular Monopole Array Antenna on a Finite Ground Plane," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 10, pp. 3406–3409, October 2005.
- [45] Yuan, H., Hirasawa, K., and Zhang, Y., "The Mutual Coupling and Diffraction Effects on the Performance of a CMA Adaptive Array," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 3, pp. 728–736, August 1998.
- [46] Furukawa, H., Kamio, Y., and Sasaoka, H., "Cochannel Interference Reduction and Path-Diversity Reception Technique Using CMA Adaptive Array Antenna in Digital Land Mobile Communications," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 2, pp. 605–616, March 2001.
- [47] Lauri, A., Colone, F., Cardinali, R., and Lombardo, P., "Analysis and Emulation of FM Radio Signals for Passive Radar," *IEEE Aerospace Conference*, Big Sky, MT, March 3–10, 2007.
- [48] IEEE Standard for Information Technology, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Std 802.11TM-1999, and following supplements and amendments, IEEE Stds 802.11aTM-1999, 802.11bTM-1999, 802.11bTM-1999/Cor 1-2001, and 802.11gTM-2003.
- [49] Bicocchi, R., Bucciarelli, T., and Melacci, P.T., "Radar Sensitivity and Resolution in Presence of Range Sidelobe Reducing Networks Designed Using Linear Programming," *Radio and Electronic Engineer*, vol. 54, pp. 224–250, June 1984.
- [50] Grant, M. and Boyd, S., CVX: MATLAB Software for Disciplined Convex Programming (web page and software), December 2009. Available at <http://cvxr.com/cvx>.
- [51] IEEE Std 802.16TM-2004, IEEE Std 802.16eTM-2005, IEEE Standard for Local and Metropolitan Area Network, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems.

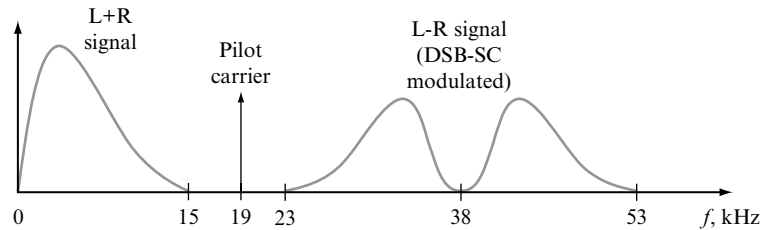
- [52] Gao, Z., Tao, R., Ma, Y., and Shao, T., "DVB-T Signal Cross-Ambiguity Functions Improvement for Passive Radar," in *Proceedings of the CIE International Conference on Radar*, Shanghai, China, October 16–19, 2006.
- [53] European Telecommunications Standard Institute, "Digital Video Broadcasting (DVB); Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television," EN 300 744, V1.1.2, 1997.
- [54] Langellotti, D., Bongioanni, C., Colone, F., and Lombardo, P., "Impact of Synchronization on the Ambiguity Function Shape for PBR Based on DVB-T Signals," *International Radar Symposium*, Vilnius, Lithuania, June 16–18, 2010.
- [55] Lombardo, P., Colone, F., Bongioanni, C., Lauri, A., and Bucciarelli, T., "PBR Activity at INFOCOM: Adaptive Processing Techniques and Experimental Results," *IEEE Radar Conference*, Rome, Italy, May 26–30, 2008.
- [56] Bongioanni, C., Colone, F., and Lombardo, P., "Performance Analysis of a Multi-Frequency FM Based Passive Bistatic Radar," *IEEE Radar Conference*, Rome, Italy, May 26–30, 2008.
- [57] Lombardo, P., Colone, F., and Bongioanni, C., "Comparison of Different Approaches for a Multi-Frequency FM Based Passive Bistatic Radar," in *Proceedings of the SET-125 Symposium on Sensors and Technology for Defence against Terrorism*, Mannheim, Germany, April 22–25, 2008.
- [58] Bongioanni, C., Colone, F., Bernardini, S., Lelli, L., Stavolo, A., and Lombardo, P., "Passive Radar Prototypes for Multifrequency Target Detection," *Signal Processing Symposium*, Jachranka, Poland, May 24–26, 2007.
- [59] Colone, F., De Leo, G., Paglione, P., Bongioanni, C., and Lombardo, P., "Direction of Arrival Estimation for Multi-Frequency FM-Based Passive Bistatic Radar," *IEEE Radar Conference*, Kansas City, MO, May 23–27, 2011.
- [60] Bongioanni, C., Colone, F., Martelli, T., D'Angeli, R., and Lombardo, P., "Exploiting Polarimetric Diversity to Mitigate the Effect of Interferences in FM-Based Passive Radar," *International Radar Symposium*, Vilnius, Germany, September 9–11, 2010.
- [61] Villano, M., Colone, F., and Lombardo, P., "Adaptive Clutter Suppression in Passive Phased Array Radar," *International Radar Symposium*, Hamburg, Germany, pp. 343–347, September 9–11, 2009.
- [62] Malanowski, M. and Kulpa, K., "Digital Beamforming for Passive Coherent Location Radar," *IEEE Radar Conference*, Rome, Italy, May 26–30, 2008.
- [63] Farina, A., *Antenna-Based Signal Processing Techniques for Radar Systems*, Artech House, Boston, MA, 1992.

17.11 | PROBLEMS

The interested reader should accomplish the following problem set in the order given using a suitable numerical simulation capability, such as the MATLAB programming environment.

1. Based on the discussion in Section 17.2.1, evaluate the computational load required by the two efficient algorithms for the 2D-CCF evaluation for a DVB-T based PBR. Assume a signal bandwidth of 8 MHz at a carrier frequency of 500 MHz (UHF band). Take the integration time to be 0.3 s and the maximum bistatic range and velocity to be 75 km and 400 m/s, respectively. Compare the obtained results and determine the most efficient algorithm to be adopted for the considered application.

2. For the same case considered in problem 1, evaluate the performance of the two sub-optimum algorithms for the 2D-CCF evaluation introduced in Section 17.2.2. Specifically, for each sub-optimum algorithm (Batches algorithm or Channelization technique), evaluate the computational load required and the SNR loss to be accepted as a function of the characteristic parameter (number of batches or number of channels, respectively).
3. Based on the results of problem 2, for each sub-optimum algorithm, select the most appropriate value for the characteristic parameter as the one which allows the best trade-off between computational load saving and SNR loss with respect to the most efficient optimum algorithm selected in problem 1. Then, compare the obtained computational load with that required by the most efficient optimum algorithm operating with a reduced number of integrated samples (so that it experiences the same SNR loss).
4. A typical FM radio signal is obtained by frequency modulating a composite stereo baseband signal whose PSD is depicted in the following figure:



The following MATLAB code performs the very basic steps for the generation of the complex envelope of an FM radio signal with duration Δ_T [sec] sampled at f_s [Hz]:

```
[audio_signal,fs_wav,n_bit] = wavread(audio_file);
audio_signal = audio_signal(1:delta_T*fs_wav,:);
audio_signal = resample(audio_signal, fs, fs_wav);
t_vect = [0:length(audio_signal)-1]'/fs;

sum_ch= 0.5 *(audio_signal(:,1)+audio_signal(:,2));
diff_ch = 0.5 *(audio_signal(:,1)-audio_signal(:,2));
stereophonic_signal = sum_ch + 0.1*cos(2*pi*19e3*t_vect) +
                        diff_ch.*cos(2*pi*38e3*t_vect);

CE_FM_radio_signal = exp(1j*2*pi*75e3*cumsum(stereophonic_signal/fs));
```

where `audio_file` is a string containing the path of a *.wav file in the 44100 Hz, 16-bit stereo format.

Use the above code to generate an FM radio signal of duration 1.5 sec, sampled at 200 kHz. Evaluate the PSD of the composite stereo baseband signal (`stereophonic_signal`), and the PSD of the resulting complex envelope (`CE_FM_radio_signal`).

5. Using the signal generated in problem 4, evaluate the corresponding 2D-CCF based on the Direct FFT algorithm described in Section 17.2.1. Set the integration time to 1 s and the maximum bistatic range and velocity to 150 km and 500 m/s, respectively. Assume a carrier frequency equal to 98 MHz. The aim of this analysis is to evaluate the exploited signal Ambiguity Function. Perform different tests using different *.wav files and compare the results in terms of achievable range resolution, sidelobes level and ambiguities for an FM-based PBR using such waveforms (e.g. see what happens for a silence transmission).
6. Using the signal generated in problem 4, build up the signal received at a surveillance channel of a PBR by exploiting the model in (17.12). Take the DNR to be 50 dB. Choose an arbitrary number of ground scatterers echoes with bistatic ranges between 5 and 50 km and CNR values between 30 dB and 45 dB. Inject 5 targets echoes with arbitrary SNR, bistatic ranges and bistatic velocities. Similarly, generate the reference signal assuming it to be a quite good replica of the transmitted signal (i.e. neglect target and clutter echoes in (17.12)) with DNR = 65 dB. Evaluate the 2D-CCF between the surveillance and the reference signals using an integration time of 1 s. Apply a CA-CFAR detection scheme and verify that most of the injected targets cannot be easily detected (unless their SNR level is high enough) due to the disturbance contributions in the surveillance channels.
7. Using the signals generated in problem 6, implement the ECA algorithm described in Section 17.3.1. Assume the ECA to operate with $K = 50$ taps. Calculate the CA between the input and the output of the cancellation filter. Then, evaluate the 2D-CCF between the reference and the surveillance signals after the disturbance cancellation stage for an integration time of 1 s. Compare the results with those obtained in problem 6.
8. Repeat the analysis in problem 7 by exploiting the ECA-B algorithm operating with $b = 10$ batches and $K = 50$ taps. Compare the results with those obtained in problems 6 and 7. Perform different tests using a different number of batches and verify that the cancellation notch widens as the number of batches increases.
9. Modify the reference signal generated in problem 6 by including a strong multipath replica (DMR = 5 dB) with arbitrary delay and phase shift (refer to Section 17.4.1). Repeat the analysis in problem 7 using the obtained reference signal and calculate the resulting CA loss. Perform different tests using different values for the delay and phase shift and verify that similar results are obtained as those reported in Figure 17-13.
10. A typical DSSS WiFi signal can be generated by chipping a random BPSK sequence at 11 MHz with a 11-chip Barker code (see Section 17.5.1). Using this approach, generate a DSSS WiFi signal of duration 0.2 s. Assume a pulsed transmission with equivalent PRF equal to 1 kHz and pulses duration equal to 0.4 ms. Evaluate the ACF for the obtained signal and verify that it is dominated by the Barker code ACF. Compute the weights of a BWN with total length 101 by using a convex programming tool (such as the MATLAB-based CVX) to solve the system in (17.26). Apply the obtained BWN to the WiFi signal and evaluate the mismatched ACF. Compare the result with that obtained for the original signal.

Appendix A: Answers to Selected Problems

Chapter 2

1. $B_F = \frac{200}{0.2} \frac{1}{10 \times 10^{-6}} = 100 \text{ MHz}$
3. $\delta R = \frac{3 \times 10^8}{2(250 \times 10^6 + (4 - 1)200 \times 10^6)} = 0.177 \text{ meters}$
5. The peak sidelobe is -27.3 dB , the peak sidelobe is -35.1 dB , the greater the TB product, the closer the peak sidelobe is to the designed-to value.
7. $N = \frac{3 \times 10^8}{2(15 \times 10^6)(0.1)} = 100 \text{ pulses}$
9. 3.259
11. (a) The weighting matrix is a diagonal matrix with diagonal elements $1 \ 1 \ 10 \ 10 \ 1 \ 10 \ 10 \ 1 \ 1$ and zeros elsewhere.
(b) $[0.1343 \quad -0.2619 \quad 0.1328 \quad 0.2619 \quad 0.1343]$

Chapter 4

2. False.
3. Use 4.13 to rewrite the weights given in 4.26 as

$$\mathbf{w}(\theta_0) = ((\mathbf{R}_\phi^T)^+ \mathbf{R}_\phi^T \otimes \mathbf{I}_N) \text{Vec}\{\mathbf{A}(\theta_0)\}$$

Use this and the definition in 4.15 to establish 4.25:

$$\begin{aligned} \mathbf{R}_e \mathbf{w}(\theta_0) &= (\mathbf{R}_\phi^T \otimes \mathbf{I}_N) ((\mathbf{R}_\phi^T)^+ \mathbf{R}_\phi^T \otimes \mathbf{I}_N) \text{Vec}\{\mathbf{A}(\theta_0)\} \\ &= (\mathbf{R}_\phi^T (\mathbf{R}_\phi^T)^+ \mathbf{R}_\phi^T \otimes \mathbf{I}_N) \text{Vec}\{\mathbf{A}(\theta_0)\} \\ &= (\mathbf{R}_\phi^T \otimes \mathbf{I}_N) \text{Vec}\{\mathbf{A}(\theta_0)\} \\ &= \mathbf{s}(\theta_0) \end{aligned}$$

4. Consider the inner product (defined as the left side of 4.2) for waveforms $s_m(t)$ and $s_n(t)$ where $m \neq n$:

$$\begin{aligned} \int_{-\infty}^{\infty} s_m(t) s_n^*(t) dt &= \frac{1}{T} \int_0^T \exp \left\{ i 2\pi \frac{m-n}{T} t \right\} dt \\ &= \frac{1}{2\pi i (m-n)} (\exp \{i 2\pi (m-n)\} - 1) \\ &= 0 \end{aligned}$$

since $e^{i 2\pi k} = 1$ for any integer k . So, the waveforms are orthogonal.

Chapter 5

2. $\mathcal{M}(\mathbf{A}) = \max_{i \neq j} |\mathbf{A}_m^H \mathbf{A}_n| = \max_{i \neq j} \frac{\sin(\pi(m-n)M/N)}{\sin(\pi(m-n)/N)}$
4. Consider a matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ with condition number 1, e.g. a random subset of a Fourier matrix. Consider a new matrix $\mathbf{B} \in \mathbb{R}^{M \times N+1}$ which is the concatenation of the matrix \mathbf{A} with an additional copy of the last column of \mathbf{A} . One can easily argue that the condition number of \mathbf{B} can be made arbitrarily close to 1 by choosing N sufficiently large for a fixed M . However, since the final two columns of \mathbf{B} are identical, $[0 \ 0 \ \dots \ -1 \ 1]^T$ is in the null space of \mathbf{B} , and thus $\mathcal{R}_2(\mathbf{B}) = 1$.
9. If \mathbf{W} is nonsingular, we can define the (unknown) variable $\mathbf{u} = \mathbf{W}\mathbf{x}$ and the modified forward-model matrix $\mathbf{M} = \mathbf{A}\mathbf{W}^{-1}$, then solve the modified problem

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmin}} \|\mathbf{u}\|_1 + \|\mathbf{M}\mathbf{u} - \mathbf{y}\|_2^2.$$

The desired solution is then recovered by the linear transformation $\hat{\mathbf{x}} = \mathbf{W}^{-1}\hat{\mathbf{u}}$.

Chapter 6

1. At low grazing angles, optical imagery provides a side-looking (or profile) view of the scene, while SAR presents a downward-looking view. Optical imagery gives a downward-looking at steep grazing angles—approaching normal incidence—while SAR imagery transitions to a profile view. Because the scene content is projected into the image plane, near-incident SAR imagery is often useless and these collection geometries are avoided.
3. Range resolution in the slant plane is always given by the familiar expression $\delta R = c/2B$. The image resolution in the ground plane is determined by the grazing angle $\delta y = \delta R / \cos \theta_{grz}$. The cross-range resolution is independent of grazing angle, and it depends only on the wavelength and angle of integration: $\delta x = \lambda/2\theta_{int}$.
7. Azimuth ambiguities will not be properly focused by the PFA. While the azimuth ambiguities alias into the scene, the rate of change of phase (or range migration) does not correspond to any true location within the scene. Thus, PFA cannot properly compensate for the ambiguous phase history.
9. In the (k_x, k_y) plane the rectangular region of support is defined by

$$k_x \in [-2k_{min} \cos \theta_{grz} \tan(\theta_{int}/2), +2k_{min} \cos \theta_{grz} \tan(\theta_{int}/2)]$$

and

$$k_y \in [2k_{min} \cos \theta_{grz}, 2k_{max} \cos(\theta_{int}/2) \cos \theta_{grz}],$$

where $k_{min} = 2\pi f_{min}/c$ and $k_{max} = 2\pi f_{max}/c$ are the minimum and maximum wavenumbers transmitted by the radar.

Chapter 7

1. Waveform bandwidth = 150 MHz
 Antenna size (horizontal) = 2 m
 Sampling interval = 0.5 m
 Integration angle (L-band) = 8.5944 deg
 Integration angle (X-band) = 0.85944 deg
 SAR baseline (L-band) = 150 m
 SAR baseline (X-band) = 15 m
3. Depth-of-focus for RDA (L-band) = 26.6667 m
 Depth-of-focus for RDA (X-band) = 266.6667 m
 For L-band, the DOF is much less than the desired swath depth of 200 m, so the variation in the PSR over downrange must be taken into account by a Stolt interpolation (RMA), or by imaging with the approximation to RDA and allowing the terms to vary over downrange. At X-band, a single PSR is good enough.
5. Area coverage rate = 20000 m²/s
 Area coverage rate = 72 km²/hr
 Achieved utility = 20000 m/s
 Potential utility = 86602540.3784 m/s

Chapter 8

1. DTED or HRE Level	0	1	2	3	4	5	6	7	8
Post Spacing (m)	928	93	31	10	3	1.1	0.5	0.25	0.125
# of Posts in 10 km ²	116	11,562	104,058	10 ⁶	1.11 × 10 ⁷	8.26 × 10 ⁷	4 × 10 ⁸	1.6 × 10 ⁹	6.4 × 10 ⁹

3. $y_0 = y_1 - h \tan \psi = 5100 - 73.6 \cdot \tan(0.361158) = 73.22$ m
5. $\hat{h}(y_1) \approx \alpha_{IF} [\phi_{ab}(y_0) - \phi_{ab}^{FE}(y_0)] = -4.3388 (444.0425 - 460.96) = 73.4016$ m
 $y_1 = 5072.2 + 73.4 \sin(20.693^\circ) = 5098.14$ m
 Error in $h = \left| \frac{73.4 - 73.6}{73.6} \right| = 0.27\%$
 Error in $y_1 = \left| \frac{5098.14 - 5100}{5100} \right| = 0.0365\%$
7. $\gamma_{noise} = \frac{1}{1 + \chi^{-1}} \Rightarrow \chi = \frac{\gamma_{noise}}{1 - \gamma_{noise}}$
 Setting $\gamma_{noise} = 0.7, 0.8$ and 0.9 gives $\chi = 2.33$ (3.68 dB), 4 (6.02 dB), and 9 (9.54 dB).

$$\begin{aligned}
 |\hat{\gamma}| &= \frac{\left| \sum_{n=0}^{N-1} a[n] b^*[n] \right|}{\sqrt{\left(\sum_{n=0}^{N-1} |a[n]|^2 \right) \left(\sum_{n=0}^{N-1} |b[n]|^2 \right)}} = \frac{\left| \sum_{n=0}^{N-1} a[n] \alpha^* a^*[n] \right|}{\sqrt{\left(\sum_{n=0}^{N-1} |a[n]|^2 \right) \left(\sum_{n=0}^{N-1} |\alpha a[n]|^2 \right)}} \\
 &= \frac{|\alpha| \left| \sum_{n=0}^{N-1} |a[n]|^2 \right|}{\sqrt{\left(\sum_{n=0}^{N-1} |a[n]|^2 \right) \left(\sum_{n=0}^{N-1} |a[n]|^2 |\alpha|^2 \right)}} = \frac{|\alpha| \sum_{n=0}^{N-1} |a[n]|^2}{|\alpha| \sqrt{\left(\sum_{n=0}^{N-1} |a[n]|^2 \right)^2}} = 1
 \end{aligned}$$

Chapter 9

1. (a) $A_{N:N-2} = A_1^N + NA_1^{N-1}(1 - A_k) + \frac{N!}{2(N-2)!}A_1^{N-2}(1 - A_1)^2$
 (b) $A_{N:N-K} = \sum_{k=0}^K \frac{N!}{k!(N-k)!}A_1^{N-k}(1 - A_1)^k$
3. If the number of elements per subarray remains 24, then increase the total number of elements to 80. Other combinations of subarray sizes and total element count are also possible.
8. $\mathbf{R}_x = |S|^2 \mathbf{v}(\theta) \mathbf{v}^H(\theta) + \sigma_n^2 \mathbf{I}$
9. $\mathbf{R}_x^{-1} = \frac{1}{\sigma_n^2} \left[\mathbf{I} - \frac{|S|^2}{\sigma_n^2 + M|S|^2} \mathbf{v}(\theta) \mathbf{v}^H(\theta) \right]$, where M is the length of the steering vector.

Chapter 11

1. Time on target: $3\text{s}/25 \text{ positions} = 120 \text{ ms}$, hence 120 pulses on target.
5. $\delta f = c/(2\Delta x)$; $\delta\theta = \lambda/(2\Delta y)$; $\delta x = c/(2\Delta f)$; $\delta y = \lambda/(2\Delta\theta)$.

Chapter 12

1. $8 \mu\text{s}$; $2 \mu\text{s/s}$
3. the RSN jammer pod; by 4 dB
5. $JSR_{SSJ} = \frac{P_j G_j d_j 4\pi R_t^2}{B_j P_t G_t \sigma n \tau}$
7. -42.4 dBmi in main lobe; -86.4 dBmi in sidelobe

Chapter 13

1. $E(z) = (\hat{x}2 + \hat{y}1) e^{-jkz}$; $\psi = \tan^{-1} \left(\frac{E_{0y}}{E_{0x}} \right) = \tan^{-1} \left(\frac{1}{2} \right) = 26.6^\circ$
3. (a) $E(z) = (\hat{x}2 + \hat{y}3e^{-j\pi/6}) e^{-jkz}$
 (b) $\psi = 57.85^\circ$; $\tau = -13.7^\circ$; $AR = \cot|\tau| = 4.1$
 (c) $P = 1.5e^{-j\pi/6} = 1.3 - j0.75$
 (d) Choose $\alpha_0 = -0.3711 \text{ (rad)}$; $\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} 2 \\ 3e^{-j\pi/6} \end{bmatrix}$
5. $E_1 = 0.9142 + j0.5$; $E_2 = 0.5 + j1.9142$
7. $E_x = 0.707$; $E_y = 1.581 e^{j1.107} = 0.707 + j1.414$; $\psi = 76.72^\circ$; $\tau = 16.85^\circ$
9. $P = 687.5 \text{ mW}$
15. $[\lambda_1 \lambda_2] = [4.1623 - 2.1623]$

$$17. \mathbf{K} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

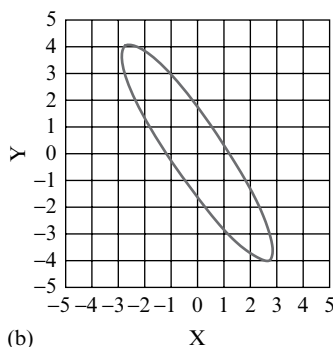
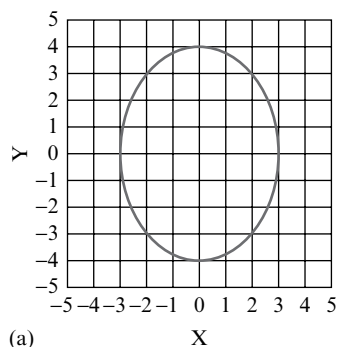
$$19. \text{Left-handed helix: } \mathbf{K} = \begin{bmatrix} 2 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 2 \end{bmatrix}; \text{Right-handed helix: } \mathbf{K} = \begin{bmatrix} 2 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

Chapter 14

1. The four steps in the Unified Framework for ATR are to (1) identify the target set, (2) select the feature set, (3) observe the feature set, and (4) test the feature set.
3. 0.3

Chapter 15

1. The off-diagonal terms influence the eigenvalues and eigenvectors, which affect the shape and rotation of the covariance ellipse. Hence, they cannot be ignored.



3. 1-D: 68.3%, 3-D: 19.9%. 3-sigma is required to get 97% containment in 3-D.
5. Measurement 2 does not gate with Track 2, since the cost of assignment is larger than the cost of coasting Track 2. The optimal answer is to assign Measurement 1 to Track 1 (cost = 5), set Measurement 2 aside for initiating a new track (cost = 20), and assign Measurement 3 to Track 2 (cost = 12) for a total assignment cost of 37.
7. Converting to linear units, the SNRs of targets 1 and 2 become 31.6 and 131.8, respectively. Hence, the difference in SNRs is still roughly 100 (in linear units) and Figure 15.5 applies. With no pulse integration, the probability of error using only SNR as a feature is found by looking at the solid ($K = 0.00$) line at the point where $R_1 = 15$ dB; the probability of error without pulse integration is roughly 0.2. If eight pulses are non-coherently integrated in this example, simply look at the dotted ($K = 0.95$) line from Figure 15.5 where $R_1 = 15$ dB; the probability of error reduce to roughly 0.01. Pulse integration separates the statistical distributions of the feature set, thereby improving the usefulness of the feature in measurement-to-track association.
9. necessity, observability, capacity, and robustness

Chapter 16

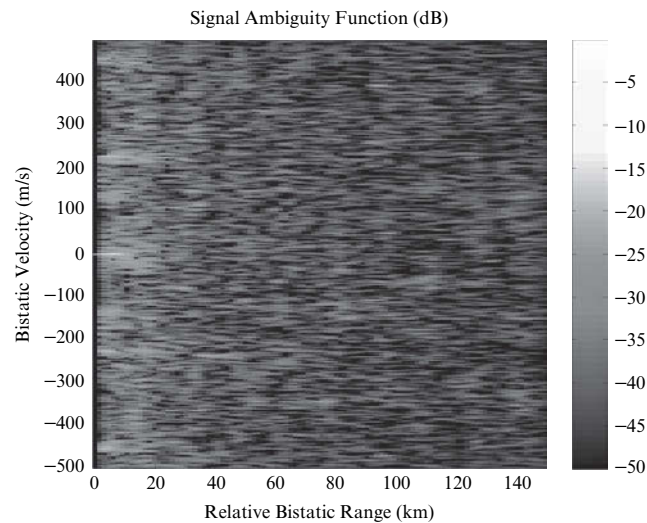
1. As RV increases, the amplitude and frequency of the torso oscillations increases.
3. The SNR loss due to phase mismatch increases as the dwell time increases.
5. The result should be similar to that of Figure 16-6.
7. The pendulum model is the most basic approximation to the true response for a dog. Better results may be obtained by increasing the number of point targets comprising the dog model, and using video data of dog motion to compute the time-varying position of each point.
9. The sinusoidal phase detector has less SNR loss than the linear phase detector.

Chapter 17

1. The Correlation-FFT requires $2.1 \cdot 10^{10}$ complex multiplications while the Direct-FFT requires $1.1 \cdot 10^{11}$ complex multiplications. The Correlation-FFT approach should be adopted in this case.
3. The Batches algorithm operating with 2000 batches yields a maximum SNR loss equal to 0.14 dB and requires $4.8 \cdot 10^9$ complex multiplications; this is 4.4 times smaller than the Correlation-FFT with full integration and 4.3 times smaller than the Correlation-FFT yielding the same SNR loss.

The Channelization technique operating with 5000 channels yields a maximum SNR loss equal to 5.35 dB and requires $1.1 \cdot 10^8$ complex multiplications; this is 188 times smaller than the Correlation-FFT with full integration and 99 times smaller than the Correlation-FFT yielding the same SNR loss.

5. An example of the results obtained for the signal generated in problem 4 is shown in the following:



9. Similar results might be obtained as those reported in Figure 17-12b.
10. Similar results might be obtained as those reported in Figure 17.

Fundamental Relations and Terminology

Major Forms of the Radar Range Equation

Peak power form, single pulse	$SNR = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 k T_0 B F L_s R^4}$		
Search form:	$\frac{P_{avg} A_e}{L_s T_0 F} \geq SNR_{min} 4\pi k \left(\frac{R^4}{\sigma} \right) \left(\frac{\Omega}{T_{fs}} \right)$		
Track form:	$\frac{P_{avg} A_e^3 k_m^2}{\lambda^4 L_s T_0 F} = \left(\frac{\pi^2}{2} \right) \left(\frac{k r N_t R^4}{\sigma \cdot \sigma_\theta^2} \right) \left(\frac{1}{\cos^5(\theta_{scan})} \right)$		
Definition of Terms:			
SNR	Signal-to-noise ratio	B	Receiver bandwidth
SNR _{min}	Minimum detectable SNR	L _s	System losses
P _t	Peak transmitted power	F	Noise figure
P _{avg}	Average transmitted power	k	Boltzmann's constant
G _t	Transmit antenna gain	Ω	Search area solid angle
G _r	Receive antenna gain	T _{fs}	Frame search time
A _e	Effective aperture	k _m	Track measurement error slope
λ	Wavelength	r	Track measurement rate
σ	Target radar cross section	N _i	Number of targets
R	Range to target	σ _θ	Track angle estimate precision (std. dev.)
T ₀	Standard temperature (270 K)	θ _{scan}	Scan angle, electronically-scanned array

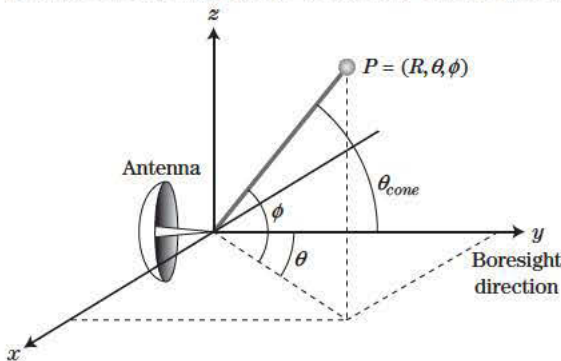
Radar Bands

Band	Frequency Range	ITU Radar Frequency
High frequency (HF)	3–30 MHz	
Very high frequency (VHF)	30–300 MHz	138–144 MHz 216–225 MHz
Ultra high frequency (UHF)	300 MHz–1 GHz	420–450 MHz 890–942 MHz
L	1–2 GHz	1.215–1.400 GHz
S	2–4 GHz	2.3–2.5 GHz 2.7–3.7 GHz
C	4–8 GHz	5.250–5.925 GHz
X	8–12 GHz	8.500–10.680 GHz
Ku (“under” K-band)	12–18 GHz	13.4–14.0 GHz 15.7–17.7 GHz
K	18–27 GHz	24.05–24.25 GHz 24.65–24.75 GHz
Ka (“above” K-band)	27–40 GHz	33.4–36.0 GHz
V	40–75 GHz	59.0–64.0 GHz
W	75–110 GHz	76.0–81.0 GHz 92.0–100.0 GHz
mm	100–300 GHz	126.0–142.0 GHz 144.0–149.0 GHz 231.0–235.0 GHz 238.0–248.0 GHz

Time Delay

A time delay of is approximately equivalent to a range of ...
1 nanosecond (ns)	0.15 meters (m)
	15 centimeters (cm)
	0.5 feet (ft)
	6 inches (in)
1 microsecond (μs)	0.15 km
	150 meters (m)
	0.1 (0.093) miles
	500 (492) feet (ft)

Definition of Azimuth, Elevation, and Cone Angles




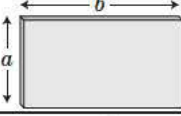
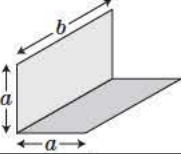
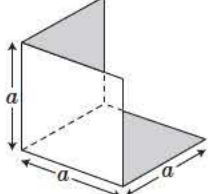
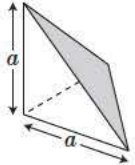
Antenna Directivity, Gain, and Beamwidth

Maximum Directivity D_{max} and Gain G
$D_{max} \approx \begin{cases} \frac{4\pi}{\theta_3 \phi_3} \alpha^2, & \theta_3, \phi_3 \text{ in radians} \\ \frac{129,600}{\pi^2 \theta_3 \phi_3} \alpha^2, & \theta_3, \phi_3 \text{ in degrees} \end{cases}$
Gain G (dB) = D_{max} (dB) – antenna losses (dB)
3 dB Beamwidth θ_3
$\theta_3 \approx \begin{cases} \frac{\alpha \lambda}{D} \text{ radians} \\ \frac{180 \alpha \lambda}{\pi D} \text{ degrees} \end{cases}$
D = aperture size
α = aperture factor
θ_3, ϕ_3 = azimuth and elevation 3 dB beamwidths

Peak Sidelobe Level, dB	Aperture Factor α
–13	0.88
–12	0.98
–25	1.05
–30	1.12
–35	1.18
–40	1.25
–45	1.30

Radar Phenomenology

Maximum RCS of Simple Shapes, $\lambda \ll \text{Object Size}$

Shape		RCS
Sphere, radius r		πr^2
Flat plate, edge lengths a and b		$4\pi(ab)^2/\lambda^2$
Dihedral, edge lengths a and b		$8\pi(ab)^2/\lambda^2$
Trihedral, square sides, edge length a		$12\pi a^4/\lambda^2$
Trihedral, triangular sides, edge length a		$4\pi a^4/3\lambda^2$

Swerling Models

Probability Density Function of RCS σ	Decorrelation	
	Scan-to-Scan	Pulse-to-Pulse
Exponential, $p_{\sigma}(\sigma) = \frac{1}{\bar{\sigma}} \exp\left[-\frac{\sigma}{\bar{\sigma}}\right]$	Case 1	Case 2
Chi-square, degree 4, $p(\sigma) = \frac{4\sigma}{\bar{\sigma}^2} \exp\left[-\frac{2\sigma}{\bar{\sigma}}\right]$	Case 3	Case 4

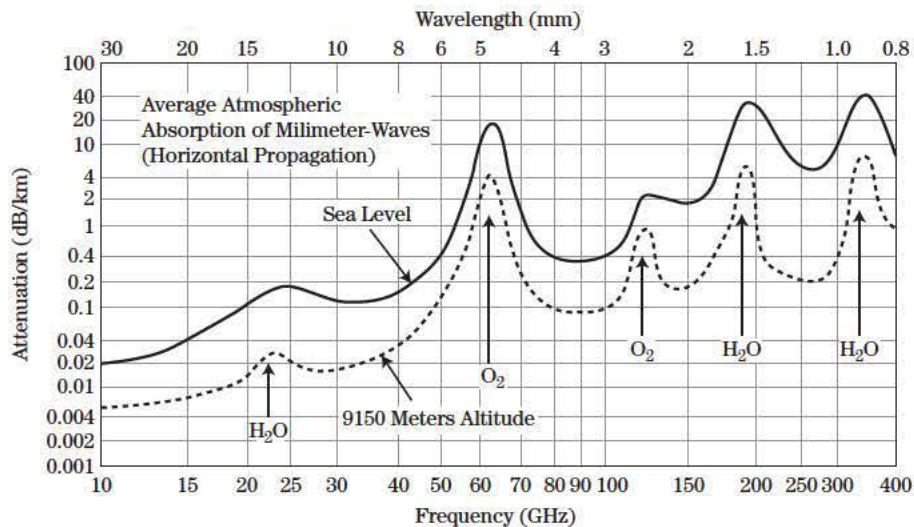
RCS Decorrelation

Variable	Required Change	Comment
Aspect angle (rad)	$\frac{c}{2Lf} = \frac{\lambda}{2L}$	L = width as viewed along radar line of sight
Frequency (Hz)	$\frac{c}{2L}$	L = depth as viewed along radar line of sight

Values of Doppler Shift

Radio frequency f		Doppler Shift f_d (Hz)		
Band	Frequency (GHz)	1 m/s	1 knot	1 mph
L	1	6.67	3.43	2.98
S	3	20.0	10.3	8.94
C	6	40.0	20.5	17.9
X	10	66.7	34.3	29.8
K _u	16	107	54.9	47.7
K _a	35	233	120	104
W	95	633	326	283

Atmospheric Attenuation



Signal Analysis and Processing

Select Fourier Transforms and Properties

Continuous Time	
$x(t)$	$X(f)$
$\begin{cases} A, & -\frac{\tau}{2} \leq t \leq \frac{\tau}{2} \\ 0, & \text{otherwise} \end{cases}$	$A\tau \frac{\sin(\pi f \tau)}{\pi f \tau} \equiv A\tau \text{sinc}(\pi f \tau)$
$\begin{cases} A \cos(2\pi f_0 t), & -\frac{\tau}{2} \leq t \leq \frac{\tau}{2} \\ 0, & \text{otherwise} \end{cases}$	$\frac{A\tau}{2} \text{sinc}[\pi(f - f_0)\tau] + \frac{A\tau}{2} \text{sinc}[\pi(f + f_0)\tau]$
$\sum_{n=-\infty}^{\infty} \delta_D(t - nT)$	$\sum_{k=-\infty}^{\infty} \delta_D(f - k \cdot PRF)$
$AB \frac{\sin(\pi Bt)}{\pi Bt} \equiv AB \text{sinc}(\pi Bt)$	$\begin{cases} A, & -\frac{B}{2} \leq f \leq \frac{B}{2} \\ 0, & \text{otherwise} \end{cases}$
$x(t - t_0)$	$e^{-j2\pi f t_0} X(f)$
$e^{+j2\pi f_0 t} x(t)$	$X(f - f_0)$
Discrete Time	
$x[n]$	$X(\hat{f})$
$Ae^{j2\pi \hat{f}_0 n}, \quad n = 0, \dots, N - 1$	$A \frac{1 - e^{-j2\pi(\hat{f} - \hat{f}_0)N}}{1 - e^{-j2\pi(\hat{f} - \hat{f}_0)}} \equiv NAe^{-j\pi(\hat{f} - \hat{f}_0)(N-1)} \text{asinc}(\hat{f} - \hat{f}_0, N)$
$A\delta[n - n_0]$	$e^{-j2\pi \hat{f} n_0}$
$\sum_{n=-\infty}^{\infty} \delta_D(t - nT)$	$\sum_{k=-\infty}^{\infty} \delta_D(f - k \cdot PRF)$
$A\hat{B} \frac{\sin[\pi \hat{B}n]}{\pi \hat{B}n} \equiv A\hat{B} \text{sinc}[\pi \hat{B}n]$	$\begin{cases} A, & \hat{f} < \hat{B} \\ \hat{B} < \hat{f} < \pi, & \text{otherwise} \end{cases}$
$x[n - n_0]$	$e^{-j2\pi \hat{f} n_0} X(\hat{f})$
$e^{+j2\pi \hat{f}_0 n} x[n]$	$X(\hat{f} - \hat{f}_0)$

Window Properties

Window	3 dB Mainlobe Width, relative to rectangular	Peak Sidelobe (dB)	Sidelobe Rolloff (dB per octave)	SNR Loss (dB)	Maximum Straddle Loss (dB)
Rectangular	1.0	-13.2	6	0	3.92
Hann	1.68	-31.5	18	-1.90	1.33
Hamming	1.50	-41.7	6	-1.44	1.68
Taylor, 35 dB, $\bar{n} = 5$	1.34	-35.2	0/6	-0.93	2.11
Taylor, 50 dB, $\bar{n} = 5$	1.52	-46.9	0/6	-1.49	1.64
Dolph-Chebyshev (50 dB equiripple)	1.54	-50.0	0	-1.54	1.61
Dolph-Chebyshev (70 dB equiripple)	1.78	-70.0	0	-2.21	1.19

Resolution

Dimension	Resolution with Matched Filter	Comments
Range (ΔR)	$\frac{c\tau}{2}$	simple pulse, length τ
	$\frac{c}{2B}$	arbitrary waveform, bandwidth B Hz
Cross-range (ΔCR)	$R\theta_3$	real beam imaging
	$\frac{\lambda R}{2vT_a} = \frac{\lambda R}{2D_{SAR}}$	synthetic aperture imaging v = platform velocity T_a = aperture time D_{SAR} = synthetic aperture size
	$\frac{\lambda}{2\gamma}$	synthetic aperture imaging γ = integration angle

Simplified Probability of Detection Estimates

Nonfluctuating target, Noncoherent integration of N samples (Albersheim's equation)	$P_D = \frac{1}{1 + e^{-B}}, \quad A = \ln \left(\frac{0.62}{P_{FA}} \right),$ $Z = \frac{SNR_{1dB} + 5 \log_{10} N}{6.2 + \frac{4.54}{\sqrt{N + 0.44}}}, \quad B = \frac{10^Z - A}{1.7 + 0.12A}$
Swerling 1 Fluctuating Target, No Noncoherent Integration	$P_D = (P_{FA})^{\frac{1}{1+SNR}}$

Miscellaneous Relations

Linear \longleftrightarrow dB Scale

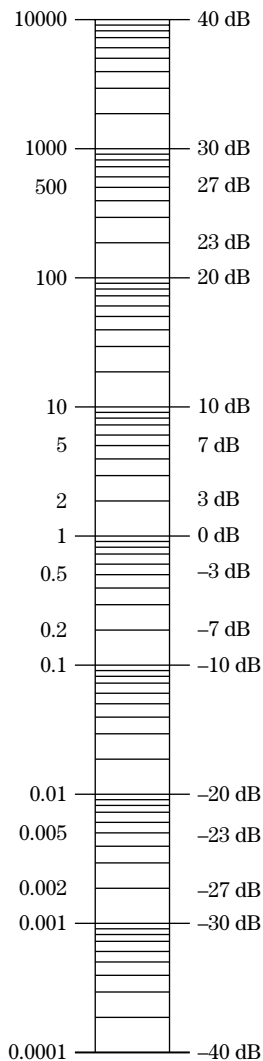


Table of Constants

Constant	Symbol	Value
Speed of light	c	2.99792458×10^8 m/s $\approx 3 \times 10^8$ m/s
Permittivity of free space	ϵ_0	8.85×10^{-12} F/m
Permeability of free space	μ_0	$4\pi \times 10^{-7}$ H/m
Impedance of free space	η	377Ω
Boltzmann's constant	k	1.38×10^{-23} J/K

Subset of AN Nomenclature Applicable to US Radar Systems

First Letter (Type of Installation)		Second Letter (Type of Equipment)		Third Letter (Purpose)	
A	Piloted aircraft	L	Countermeasures	D	Direction finder, reconnaissance, or surveillance
F	Fixed Ground	P	Radar	G	Fire control or searchlight directing
M	Ground, mobile (installed as operating unit in a vehicle which has no function other than transporting the equipment)	Y	Signal/data processing	K	Computing
P	Pack or portable (animal or man)			N	Navigational aids (including altimeter, beacons, compasses, racons, depth sounding, approach, and landing)
S	Water surface craft			Q	Special, or combination of purposes
T	Ground, transportable			R	Receiving, passive detecting
U	Ground utility			S	Detecting or range and bearing, search
V	Ground, vehicular (installed in vehicle designed for functions other than carrying electronic equipment, etc., such as tanks)			Y	Surveillance (search, detect, and multiple target tracking) and control (both fire control and air control)

Supplements and errata available at
www.scitechpub.com/pomr2

Index

Note: Page numbers followed by “*f*” and “*t*” indicate figures and tables respectively.

A

- AACER. *See* Affordable Adaptive Conformal ESA Radar (AACER) program
- Absolute elevation estimation, InSAR, 374–375
- ACF. *See* Autocorrelation function (ACF)
- ACN. *See* Additive colored noise (ACN)
- Active electronically scanned array (AESA), 87, 111–112, 112*f*, 124
- Adaptive antenna array, for passive radar, 808–813, 810*f*–812*f*, 814*f*
- Adaptive cancellation filter, 756–757
- Adaptive digital beamforming (ADBF), 1, 7, 10–11
- Adaptive feedback, SAR, 647
- Adaptive jammer cancellation. *See* Jammer cancellation, adaptive algorithms for
- Adaptive MIMO radar, 109–113
 - dynamic calibration, 111–112, 112*f*
 - transmit-independent channel estimation, 110–111, 110*f*, 111*f*
- ADBF. *See* Adaptive digital beamforming (ADBF)
- ADC. *See* Analog-to-digital converter (ADC)
- Additional sidelobes reduction filter (ASRF), 778–780, 780*f*
- Additive colored noise (ACN)
 - example, 92–95, 93*f*, 94*f*
 - jamming, 110
 - optimum MIMO waveform design for, 89–95, 89*f*, 90*f*, 93*f*, 94*f*
 - receivers for, 90, 90*f*
- Additive Gaussian colored noise (AGCN), 90
 - multichannel radar for, 89*f*
- Additive noise, 242
- Additive white Gaussian noise (AWGN), 287
- Advanced SAR (ASAR),
 - ENVISAT, 384
- AESA. *See* Active electronically scanned array (AESA)
- AF. *See* Ambiguity function (AF)
- Affordable Adaptive Conformal ESA Radar (AACER) program, 707
- AGC. *See* Automatic gain control (AGC)
- AGCN. *See* Additive Gaussian colored noise (AGCN)
- Airborne surveillance radars, colored
 - transmission for, 514, 514*f*
- Airborne systems, 385–386
 - spaceborne vs., 360–361
- Aircraft skin return, 543
- Alignment, DBF, 407
- Along-track interferometry (ATI), 386, 389–392, 391*f*
 - applications, 390–391, 391*f*
- Along-track sampling rate, SAR systems, 138–139, 229–231, 230*f*
- Alpha-beta filter, 675
- Alternate CFAR detection, 572–574, 573*f*
- Alternative hypothesis, radar detection, 2, 3
- AM. *See* Amplitude modulation (AM)
- Ambiguity function (AF), 153–155
 - PBR, 742–743
- Ambiguity-to-signal ratio (ASR), 242–243
- AMP. *See* Approximate message passing (AMP)
- Amplitude modulation (AM), 538
 - radio, 740
- Amplitude taper, 48
 - Fourier transform, 51
 - range sidelobes suppression and, 38–39, 38*f*
- AMR. *See* Lynx Advanced Multi-Channel Radar (AMR)
- Analog-to-digital converter (ADC), 7, 404, 464, 539
 - beat frequency and, 30, 34
 - for CW pulse, 60
 - ENOB, 26
 - hardware, limits of, 184
 - processing bandwidth, 37
 - sampling rate, 26, 37, 41–42, 226–227
- Angle alignment correlation, 577, 577*f*
- Angle ambiguity function, for MIMO radar, 141
- Angle break-locks (track breaks), 544
- Angle estimation, STAP, 490
- Angle false targets, 543
- Angle gating, 580–581, 581*f*
- Angle-of-arrival (AOA), 97, 104
 - of pulse, 538
- Angle-on-jam, 580
- Angle track deception techniques, 544–545
- Angle variance, signal processor-based EP and, 574
- Angular anisotropy, 187
- AN/PPS-5 Ground Surveillance Radar, 705
- Antenna
 - area, stripmap SAR, 322–323
 - auxiliary, 555
 - based EP, 554–561
 - low cross polarization antenna, 560–561
 - low sidelobes, 555
 - MLC, 557–558
 - monopulse angle measurement, 559–560, 560*f*
 - narrow beamwidth, 558
 - passive conical scanning, 558–559, 559*f*
 - random beam scanning, 558
 - sector blanking, 555
 - SLB, 555–556, 556*f*
 - SLC, 556–557
 - patterns, jammer cancellation, 434
 - sidelobe canceller, 556
 - systems, active
 - space-time coding for. *See* Space-time coding, for active antenna systems
- Antiradiation missile (ARM), 535
- AOA. *See* Angle-of-arrival (AOA)
- Appendage-to-torso ratio (ATR), 721
- Approximate message passing (AMP), 169, 179–180
- ARM. *See* Antiradiation missile (ARM)
- Array, MIMO virtual, 122–123
- Array factor (AF), of MIMO radar, 131, 132*f*

ASAR. *See* Advanced SAR (ASAR)
 Aspect angle, 618
 ASR. *See* Ambiguity-to-signal ratio (ASR)
 ASRF. *See* Additional sidelobes reduction filter (ASRF)
 Assignment algorithms, 674–675
 A160T Hummingbird UAS, 707
 ATI. *See* Along-track interferometry (ATI)
 Atmospheric effects, InSAR, 378, 379
 ATR. *See* Appendage-to-torso ratio (ATR); Automatic target recognition (ATR)
 Autocorrelation function (ACF), 776, 777f
 Autofocus techniques, SAR, 251–253
 phase difference autofocus, 252–253
 phase gradient autofocus, 253, 254f
 Automatic gain control (AGC), 542, 544
 capture, 569
 deception, 569
 Automatic target recognition (ATR), 12
 acronyms, 632–633
 common metrics
 confusion matrices, 635, 635t
 receiver operating characteristic (ROC) curves, 634–635
 framework, steps for, 13
 HRRPs. *See* High-resolution range profiles (HRRPs)
 ISAR. *See* Inverse synthetic aperture radar (ISAR)
 key points, 632
 organization, 632
 overview, 631–633
 passive radar, 656–658. *See also* Passive radar ATR
 performance prediction
 in Bayesian framework, 636–637
 comparison of algorithms via Wald sequential test, 637
 in Neyman-Pearson framework, 635–636
 tools in design process, 637–638
 SAR. *See* Synthetic aperture radar (SAR)
 unified framework for, 633–634
 challenges, 634
 identifying target set, 633
 observing feature set, 633
 selecting feature set, 633
 testing feature set, 634
 Auxiliary angle (γ), 596, 597, 598
 Auxiliary antennas, 555
 Axial ratio, defined, 596
 Azimuth angle, 236
 Azimuth dechirp, 271–276, 273f, 274f–276f

B

Back-scatter alignment (BSA), 601–602
 BAE Systems INSYTE, 411, 411f

Bandwidth
 composite waveform, 40
 filter, 34
 waveform, 26–27
 Barker sidelobes reduction (BWN)
 weighting network for, 777–778, 778t
 Barrage noise (BN), 540, 541
 Baseband signal, 41
 analytical representation of, 30
 spectrum of, 31–32, 32f
 Baseline decorrelation, InSAR, 377–378
 Basis Pursuit De-Noising (BPDN), 159, 168–170, 170f, 175
 Basket-weave plots, 233
 Batches algorithm, suboptimum
 implementations of 2D-CCF and, 751–753, 752f, 753f
 Bayesian approaches, 178–180, 180f
 Bayesian compressive sensing (BCS)
 algorithm, 184
 Bayesian framework, ATR
 performance prediction, 636–637
 Bayesian network, SAR, 650
 BBN. *See* Blinking barrage noise (BBN)
 BCS algorithm. *See* Bayesian compressive sensing (BCS) algorithm
 Beamformer architectures, 435–441
 beamspace adaptive canceller, 440, 441f
 sidelobe blanker, 435–437, 435f–437f
 sidelobe canceller, 437–440, 438f–440f
 subarray space adaptive cancellation, 441
 Beamforming, 455, 459–464
 adaptive, 95
 spatial, MIMO radar, 130
 Beampatterns
 MIMO radar, 142
 as pattern of array antenna, 132–133, 133t
 Beamspace adaptive canceller, 440, 441f
 Beat frequency, 30, 30f, 33
 Doppler shift and, 39, 39f
 filter bandwidth and, 34
 Belief propagation (BP), 179–180, 180f
 Bernoulli indicator variable, 179
 Bidimensional frequency coding, 507–508, 508f
 Binary-phase coding, pseudo-random, 565
 Biphas-coded waveforms, 70
 Biphas-to-quadruphase (BTQ)
 transformation, 70–72, 71f, 72f, 73
 Bistatic polarimetric SAR, 640
 Bistatic radar
 passive. *See* Passive bistatic radar (PBR)
 topology, 6
 Blanking loss, 556
 Blinking barrage noise (BBN), 541
 Blinking spot noise (BSN), 541
 Block diagram, STAP, 490–491, 491f

Block sparsity, 181
 BN. *See* Barrage noise (BN)
 Born approximation, 189
 Boulic-Thalmann model, 713–716, 713f, 715f, 716f, 727, 728, 729
 time-varying ranges computation with, 716–718, 717t
 BP. *See* Belief propagation (BP)
 BPDN. *See* Basis Pursuit De-Noising (BPDN)
 Broadcast transmitters, 739
 BSA. *See* Back-scatter alignment (BSA)
 BSN. *See* Blinking spot noise (BSN)
 BTQ transformation. *See* Biphas-to-quadruphase (BTQ) transformation
 Burnthrough waveforms, 563–564
 BWN. *See* Barker sidelobes reduction (BWN)

C

CA. *See* Clutter attenuation (CA)
 CA-CFAR. *See* Cell-averaging constant false alarm rate (CA-CFAR)
 Calgis, Inc., 385
 Calibration, DBF, 407
 California Department of Conservation, 385
 Canadian Centre for Remote Sensing (CCRS), 384, 385
 Cancellation ratio (CR), 433, 557
 Carnegie Mellon University (CMU) Motion Research Laboratory, 711
 Cascaded integrator-comb (CIC) filter, 753
 Cauchy-Schwarz theorem, 92
 C band SAR, 385
 CC-CFAR. *See* Censored cell CFAR (CC-CFAR)
 CCD. *See* Coherent change detection (CCD)
 CCRS. *See* Canadian Centre for Remote Sensing (CCRS)
 CDM. *See* Code division multiplexing (CDM)
 CEC. *See* Cooperative Engagement Capability (CEC)
 Cell-averaging CFAR, 573
 Cell-averaging constant false alarm rate (CA-CFAR), 87
 Cell under test (CUT), 572, 573
 Censored cell CFAR (CC-CFAR), 573–574
 Center reference point (CRP), 324, 325
 Centralized fusion. *See* Measurement-level fusion
 Centralized linear integration, 795, 795f, 796
 CFAR. *See* Constant false alarm rate (CFAR)
 Chaff, 544–545, 571
 Channelization technique, suboptimum
 implementations of 2D-CCF and, 753–755

- Channel matrix, MIMO, 125–126
- Chernoff information, 635
- Chirp scaling algorithm (CSA), 315–318, 317*f*
- Cholesky decomposition, 160
- Cholesky factorization, 432
- Chopped repeater, 540
- CIC filter. *See* Cascaded integrator-comb (CIC) filter
- Circular polarization, 595, 595*f*
 - Poincaré sphere and, 600, 600*f*
- Circulating code, 506–507, 507*f*
- Circulating pulse, 504–505, 504*f*, 505*f*, 506*f*
- CLEAN algorithm, 177
- Cluster discrimination, 645
- Cluster tracking, 686
- Clutter
 - ground, space-time properties of, 472–474
 - heterogeneous, 492
 - nonstationary, 492
- Clutter attenuation (CA)
 - loss of ECA, 767–768, 768*f*
- Clutter rejection, 517–520
 - space domain, 518–520, 518*f*, 519*f*
 - time domain, 517, 517*f*
- Clutter-to-noise ratio (CNR), 242, 471
- CMA. *See* Constant modulus algorithm (CMA)
- CNR. *See* Clutter-to-noise ratio (CNR)
- Code division multiplexing (CDM), 135
- Coefficients, in generation of NLFM waveforms, 55, 55*t*
- Coherence, defined, 535
- Coherence and statistical errors, InSAR, 376–381
 - atmospheric and weather effects, 378, 379
 - baseline decorrelation, 377–378
 - coherence and interferometric system design, 379–380–381
 - interferogram phase statistics, 379–380, 380*f*
 - multiple looks, 379–380
 - temporal decorrelation, 378, 379*f*
 - thermal and quantization noise, 376–377
- Coherent change detection (CCD), 388–389, 389*f*
- Coherent cover pulse, 541
- Coherent detector, 30
- Coherent exploitation techniques, 386–392
 - along-track interferometry (ATI), 386, 389–392, 391*f*
 - coherent change detection (CCD), 388–389, 389*f*
 - terrain motion mapping, 386–388, 387*f*, 388*f*
- Coherent integration, non-coherent vs., 512, 513*f*
- Coherent jammers, 536*f*, 537–540
 - computation example, 551–552, 552*f*
 - DRFM for, 539–540, 539*f*
 - formulas, 549–551
- Coherent masking techniques, 542
- Coherent noise, 542
- Coherent oscillator (COHO), 620
- Coherent processing interval (CPI), 155, 156, 563, 719, 748
- Coherent repeater, 540
- Coherent waveform, multiple pulse, 545
- COHO. *See* Coherent oscillator (COHO)
- Colored noise matched filter, 90
- Colored space-time exploration, 500–514
 - colored transmission (space-time coding). *See* Colored transmission
 - colored transmission trade-offs and applications, 513–514, 514*f*
 - digital beamforming. *See* Digital beamforming (DBF)
 - diversity gain, 509, 512–513, 513*f*
 - target coherence, 509–511, 510*f*, 511*f*
- Colored transmission, 502–508, 502*f*. *See also* Space-time coding
 - bidimensional frequency coding, 507–508, 508*f*
 - circulating code, 506–507, 507*f*
 - circulating pulse, 504–505, 504*f*, 505*f*, 506*f*
 - fast scanning (intra-pulse scanning), 505–506, 506*f*
 - optimum processing, 503, 503*f*
 - principles, 502–504, 503*f*
 - trade-offs, applications and, 513–514, 514*f*
- Commercial off-the-shelf (COTS) components, 409
- Composite waveform, 42
 - bandwidth of, 40
- Compressed sensing (CS)
 - ambiguity function, 153–155
 - belief propagation, 181
 - issues associated with, 166
 - Kruskal rank, 162–163
 - linear model, 150–151
 - mutual coherence, 165–166
 - overview, 147, 150
 - radar applications, 183–195
 - radio frequency tomography, 152–153, 152*f*
 - restricted isometry property, 163–165
- Compressive sampling matching pursuits (CoSaMP), 177–178
- Confusion matrices, ATR and, 635, 635*t*
- Conical scan (CONSCAN) radar, 558–559, 559*f*
- Conical-scan-on-receive-only (COSRO), 559
- Conopulse, 560
- CONSCAN. *See* Conical scan (CONSCAN) radar
- Constant false alarm rate (CFAR), 479, 638, 644, 740
 - algorithm, 7
 - detection, 542, 572–574, 573*f*
- Constant gamma model, 221
- Constant modulus, NLFM waveforms and, 108–109, 109*f*
- Constant modulus algorithm (CMA), 769, 770*f*
 - for reference signal cleaning using analog modulation, 767–771, 768*f*, 770*f*, 771*t*, 772*f*
- Constrained optimization, jammer cancellation, 425–429
 - derivative constraints, 428
 - eigenvector constraints, 428–429
 - GSC, 427–428, 428*f*
 - LCMV, 426–427, 427*f*
- Continuous wave (CW) pulse, 59
 - ADC for, 60
 - matched filter response for, 64, 64*f*
- Contourlet transform, 655
- Contrast ratio (CR), 241
 - values of, 242
- Cooperative bistatic radar, 6
- Cooperative blinking noise (CBN), 545
- Cooperative Engagement Capability (CEC), 692
- Coordinate systems, RMA, 306–307
- Copol max, 605
- Copol saddle, 605
- Correlation FFT, 2D-CCF implementations and, 749–750, 749*f*, 750*t*
- Correlation filter, SAR, 651
- Correlation processing, stretch processing vs., 36–37, 37*t*
- Corridor chaff, 545
- CoSaMP. *See* Compressive sampling matching pursuits (CoSaMP)
- COSRO. *See* Conical-scan-on-receive-only (COSRO)
- Cost functions, 674
- Covariance matrix, 669
- Covariances, 671, 671*f*
- CPI. *See* Coherent processing interval (CPI)
- CR. *See* Contrast ratio (CR)
- Cramer-Rao Lower Bound (CRLB), 696
- Critical array extent (D_c), 510, 511*f*
- Critical bandwidth (Δf_c), defined, 510
- CRLB. *See* Cramer-Rao Lower Bound (CRLB)
- Cross-eye (XEYE), 544
- Cross-pol, 544

- Cross-polarization (XPOL), 544, 560, 561
- Cross-range resolution, for SAR, 234
- CRP. *See* Center reference point (CRP)
- CSA. *See* Chirp scaling algorithm (CSA)
- Cumulative distribution functions, jammer cancellation, 434–435, 435*f*
- CUT. *See* Cell under test (CUT)
- CVXMATLAB-based modeling system, 778
- CW pulse. *See* Continuous wave (CW) pulse
- C/X-SAR, 385
- D**
- DAB transmitters. *See* Digital audio broadcasting (DAB) transmitters
- DAC. *See* Digital-to-analog converter (DAC)
- Dantzig selector, 168
- Data editing technique, signal processor-based EP and, 574–575
- Data least squares (DLS), 192
- Data processor-based EP, 576–581
 - angle alignment correlation, 577, 577*f*
 - angle gating, 580–581, 581*f*
 - range-Doppler track comparison, 578–579
- RCS
 - statistics, 577
 - step detection, 578
 - track filter acceleration limit, 579–580
 - track-on-jam, 580
- DAWGs. *See* Digital arbitrary waveform generators (DAWGs)
- DBF. *See* Digital beamforming (DBF)
- DBS. *See* Doppler beam sharpening (DBS)
- DCCF. *See* Distance classifier correlation filter (DCCF)
- 2D-CCF. *See* Two-dimensional cross-correlation function (2D-CCF)
- Decentralized fusion. *See* Track-level fusion
- Decentralized integration L/Nch (DEC), 795–796
- Deception EA techniques, 533, 534, 542–545
 - angle track deception, 544–545
 - false targets, 542–543, 543*f*
 - range track deception, 543–544
 - velocity track deception, 543–544
- Degrees of freedom (DoF), 3–4, 14, 454
 - clutter, 487
 - computational loading, 484
 - Doppler, 485
 - EFA, 488
 - eigenspectrum, 471
 - GSC, 482
 - JDL, 487, 488
 - spatial and temporal, 454, 487
 - transmit/receive, 90
- Delay-Doppler cells, 195, 195*f*
- Dempster-Shaffer algorithm, 650
- DEMs. *See* Digital elevation maps (DEMs)
- De-noising, 645
- Dense MIMO array, 140
- Depression angle, with pairs of range measurements, 345–347, 346*f*
- Deramping, 30
- Deramp reception, 226–229
- Derivative constraints, 428
- Detected video, 537
- Detection, 2, 3–4
 - sensitivity, 552
- DFT. *See* Discrete Fourier transform (DFT)
- Dicke-Fix. *See* Wideband limiting
- Digital arbitrary waveform generators (DAWGs), 87
- Digital audio broadcasting (DAB) transmitters, 739
- Digital beamforming (DBF), 401, 500–501, 500*f*, 557
 - analog-to-digital converters, 404
 - architecture, 405
 - benefits, 401, 414
 - digitally restereed antenna pattern, 415–419
 - element-level, 404*f*
 - FIR filters, 406, 406*f*
 - fundamentals, 404–406
 - implementation challenges, 406–411
 - alignment and calibration, 407
 - computational load, 410
 - cost, 406
 - data movement, 407
 - higher frequencies, 407, 407*f*
 - input-output (I/O) throughput, 409, 409*f*
 - interference, 406
 - packaging, 406
 - power, 406
 - receivers, 407–409
 - reliability, 407
 - single-channel availability, 408, 409*f*
 - I/Q samples, 404
 - limitations, 501
 - principle and objectives, 500–501
 - processor, operation of, 405, 406*f*
 - subarrays, 411–414. *See also* Subarrays, digital beamforming (DBF)
 - time-delay steering, 405–406
 - WFG, 404–405
 - wideband array, 405
- Digital elevation maps (DEMs), 10
- Digital height model (DHM), 342
- Digitally restereed antenna pattern, 415–419
 - for irregular subarray, 417, 417*f*
 - for overlapped subarray, 417–418, 417*f*–418*f*
- for regular subarray, 416, 416*f*
- Digital radio mondiale (DRM), 740
- Digital RF memory (DRFM), 539, 539*f*, 542
- Digital sinc function, 35
- Digital surface model (DSM), 342
- Digital Terrain Elevation Data (DTED), 342–344, 342*t*
- Digital terrain model (DTM), 342–344
 - algorithm, 349–351, 351*f*
 - DTED, 342–344
 - HRE, 344
- Digital-to-analog converter (DAC), 539
- Digital transmissions, reference signal reconstruction for, 761–775, 774*f*
- Digital video broadcasting-terrestrial (DVB-T) transmitters, 739
- Direct digital synthesis (DDS), 536
- Direct FFT, 2D-CCF implementations and, 749*f*, 750–751, 751*t*
- Direct signal and multipath/clutter cancellation techniques, in PBR, 755–766
 - ECA-B approach, 760–761, 760*f*
 - ECA-B&S algorithm, 761–766, 762*f*, 763*t*, 764*f*–765*f*
 - extensive cancellation algorithm, 758–760, 759*f*
- Direct signal-to-noise ratio (DNR), 767
- Dirichlet function, 35
- Discrepancy principle, Morozov, 190
- Discrete Fourier transform (DFT), 27, 30, 34–35, 36, 63, 359, 749
 - defined, 36
 - inverse, 52
 - K-length, 44
 - SR algorithms, 166–167
 - straddle loss, 63
- Discrete time Fourier transform (DTFT) output of, 61
- SF waveform, 62–63
- Discrete wavelet transform (DWT), 167
- Dismount detection. *See* Human-dismount detection
- Distance classifier correlation filter (DCCF) SAR, 651
- Distributed target contrast ratio (DTCR) defined, 241
 - quantities contributing to, 241*t*
- Diversity gain, 509, 512–513, 513*f*
- DL PUSC, 782
- DLS. *See* Data least squares (DLS)
- DN. *See* Doppler noise (DN)
- DNR. *See* Direct signal-to-noise ratio (DNR)
- DoF. *See* Degrees of freedom (DoF)
- Doppler beam sharpening (DBS), 261, 269–270, 269*f*, 271*f*
 - azimuth dechirp, 271–276

notes on, 285
 RMC. *See* Range migration compensation (RMC)
 for SAR, 9–10
 Doppler bin masking, 542
 Doppler estimation, STAP, 490
 Doppler false targets, 542
 Doppler frequency, 156, 231
 Doppler noise (DN), 542
 Doppler resolution waveform, 565–566
 Doppler shift, 57, 57*f*, 58*f*
 and beat frequency, 39, 39*f*
 impact on SF waveform, 68–69
 and mismatched filters, 78, 79*f*
 Doppler-walk, 564
 Dots, GMTI radar detections, 2–3, 13
 Double sliding window CFAR approach, 573
 DRFM. *See* Digital RF memory (DRFM)
 DRM. *See* Digital radio mondiale (DRM)
 Drop phase, in pull-off sequence, 544
 DTCR. *See* Distributed target contrast ratio (DTCR)
 DTFT. *See* Discrete time Fourier transform (DTFT)
 Durability, EP, 554
 DVB-T signals, 2D-CCF shape control for, 786–791, 787*f*–790*f*, 787*t*
 DVB-T transmitters. *See* Digital video broadcasting-terrestrial (DVB-T) transmitters
 Dwell phase, in pull-off sequence, 544
 DWT. *See* Discrete wavelet transform (DWT)

E

EA. *See* Electronic attack (EA)
 Earth Remote Sensing Data Analysis Center (ERSDAC), 384–385
 Eaton Corporation, 705
 ECA. *See* Extensive cancellation algorithm (ECA)
 ECA-B approach, 760–761, 760*f*
 ECA-B&S algorithm, 761–766, 762*f*, 763*t*, 764*f*–765*f*
 Edge detection, 655
 EFA. *See* Extended factored algorithm (EFA)
 Effective radiated power (ERP), 12, 547
 EH-60 Black Hawk, 707
 Eigenvector constraints, 428–429
 EIRP. *See* Equivalent Isotropic Radiated Power (EIRP)
 EKF. *See* Extended Kalman filter (EKF)
 Electromagnetic compatibility (EMC), 566
 Electromagnetic interference (EMI), 7, 553
 Electronically scanned array (ESA), 557
 Electronic attack (EA), 12
 deception. *See* Deception EA techniques

 defined, 529
 jammer in. *See* Jammers
 masking. *See* Masking EA techniques
 overview, 533–535, 534*f*
 techniques, summary, 582*t*
 Electronic counter-counter measures (ECCM), 529
 Electronic counter measures (ECM), 529
 Electronic protection (EP), 12, 529–584
 abbreviations, 531–533
 antenna-based. *See* Antenna, based EP
 data processor-based. *See* Data processor-based EP
 EA. *See* Electronic attack (EA)
 EW-related formulas. *See* Electronic warfare (EW)
 exciter-based. *See* Exciter-based EP
 key points, 530
 notation, 530–531
 organization, 530
 overview, 529–533, 553–554
 receiver-based. *See* Receiver-based EP
 signal processor-based. *See* Signal processor-based EP
 techniques, summary of, 583*t*
 transmitter-based. *See* Transmitter-based EP
 Electronic support (ES), 529
 Electronic support measures (ESM), 529
 Electronic warfare (EW), 529
 band designations, 535*t*
 related formulas, 545–553
 coherent jammer. *See* Coherent jammers
 jammer received power. *See* Jammers, received power
 keeping track of losses, 546–547
 noise jammer. *See* Noise, jammers
 signal and thermal noise, 545–546
 Electro-optical/infrared (EO/IR) focal plane arrays, 631
 Elevation measurement, InSAR, 344–359
 DEM algorithm, 349–351, 351*f*
 depression angle with pairs of range measurements, 345–347, 346*f*
 layover and foreshortening, 355–358, 356*f*–357*f*, 358*f*
 reference profile, 355–356
 spatial baseline effect, 354
 3-D SAR, 358–359
 2-D SAR images, 344–345, 345*f*
 variations from phase measurements, 347–349
 wrapped phase, 351–354, 353*f*
 Elliptical polarization, 596–598, 596*f*
 Poincaré sphere and, 600–601, 600*f*
 Ellipticity angle (τ), 596, 598

EMCON. *See* Emission control (EMCON)
 EMI. *See* Electromagnetic interference (EMI)
 Emission control (EMCON), 562
 Environmental Research Institute of Michigan (ERIM), 385
 ENVISAT satellite, 384
 EO/IR. *See* Electro-optical/infrared (EO/IR) focal plane arrays
 EP. *See* Electronic protection (EP)
 Equivalent Isotropic Radiated Power (EIRP), 740
 ERIM. *See* Environmental Research Institute of Michigan (ERIM)
 ERIM/NAWC P-3 Ultra-Wideband SAR, 330
 ERP. *See* Effective radiated power (ERP)
 Error sources, InSAR
 coherence and statistical, 376–381
 atmospheric and weather effects, 378, 379
 baseline decorrelation, 377–378
 coherence and interferometric system design, 379–380–381
 interferogram phase statistics, 379–380, 380*f*
 multiple looks, 379–380
 temporal decorrelation, 378, 379*f*
 thermal and quantization noise, 376–377
 systematic, 381–382
 sensor geometry errors, 381
 ERSDAC. *See* Earth Remote Sensing Data Analysis Center (ERSDAC)
 ERS-1/ERS-2. *See* European Remote Sensing (ERS) 1 and 2 satellite
 ES. *See* Electronic support (ES)
 ESA. *See* European Space Agency (ESA)
 Escort support jammer (ESJ), 534
 ESJ. *See* Escort support jammer (ESJ)
 ESPRIT. *See* Estimation of signal parameters via rotational invariance techniques (ESPRIT)
 Estimation of signal parameters via rotational invariance techniques (ESPRIT), 612, 613–614
 Euclidean distance, 151
 Euler decomposition, ISAR, 654
 European Remote Sensing (ERS) 1 and 2 satellite, 384
 European Space Agency (ESA), 384
 Evolutionary programming, 648
 EW. *See* Electronic warfare (EW)
 Exciter-based EP, 562–567
 burnthrough waveforms, 563–564
 Doppler resolution waveform, 565–566

- Exciter-based EP (*cont.*)
 frequency variation, 566
 multiple-pulse waveform, 563
 multiple simultaneous frequency radiation, 566–567
 PRF variation, 567
 range resolution
 narrow-pulse, 564
 pulse compression, 564–565
 wide-pulse waveform, 562–563
- Expendable decoy, 545
- Extended clutter (chaff), 571
- Extended factored algorithm (EFA), 485–486, 485*f*, 487, 488
- Extended Kalman filter (EKF), 13, 675
- Extensive cancellation algorithm (ECA), 758–760, 759*f*
- F**
- False targets, 542–543, 543*f*
- Faraday rotation
 effect of, 590
 of polarization plane, 589
- Fast Bayesian matching pursuits (FBMP)
 approach, 184
- Fast Fourier transform (FFT), 36, 63, 172, 234, 236, 645, 720
 2D-CCF implementations and correlation FFT, 749–750, 749*f*, 750*t*
 direct FFT, 749*f*, 750–751, 751*t*
- Fast Iterative Shrinkage-Thresholding Algorithm (FISTA), 169, 174
- Fast scanning (intra-pulse scanning), 505–506, 506*f*
- Fast-time, defined, 716–720
- FBMP approach. *See* Fast Bayesian matching pursuits (FBMP) approach
- FDM. *See* Frequency division multiplexing (FDM)
- Feature-assisted tracking, 678–682, 681*f*
- FFT. *See* Fast Fourier transform (FFT)
- Field programmable gate arrays (FPGA), 409, 410
- Finite impulse response (FIR) filters, 42, 406
- FIR filters. *See* Finite impulse response (FIR) filters
- FISTA. *See* Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)
- FM. *See* Frequency modulation (FM)
- FOA. *See* Focus of attention (FOA) module
- Focal underdetermined system solver (FOCUSS) algorithm, 176
- Focus of attention (FOA) module, 638
- FOCUSS algorithm. *See* Focal underdetermined system solver (FOCUSS) algorithm
- Foliage Penetration, Reconnaissance, Surveillance, Tracking and Engagement Radar (FORESTER) program, 707
- FORESTER. *See* Foliage Penetration, Reconnaissance, Surveillance, Tracking and Engagement Radar (FORESTER) program
- Forward-looking array (FLA), 473
- Forward-scatter alignment (FSA), 601, 602
- Fourier analysis, 61
- Fourier ribbon, 235–236, 236*f*
- Fourier transform, 7, 31, 645
 amplitude taper, 51
 inverse, 50
 matched filtering, 270, 287
 RDA, 295, 301, 303–304, 306
 RMA, 307
 RMC procedure, 280
 RSA, 290–292
 SAR imaging, 214–215, 216*t*
 of shifted delta function, 218
 squinted stripmap, 326
 Stolt approximation, 313
 time-frequency scaling property of, 227
 two dimensional, 448
 wideband phase errors, 249
- FPGA. *See* Field programmable gate arrays (FPGA)
- Fractional Doppler shift, 48
- Fraility, EP, 554
- Free-space attenuation, 546
- Free-space loss, 546
- Frequency
 agility, 566
 discriminator, 537
 diversity, 566
 RMA, 306–307
 variation, 566
- Frequency division multiplexing (FDM), 135
- Frequency domain (FD)
 analysis, with complex image, 644
 example, 46–47, 46*f*, 47*f*, 48*f*
 implementation steps, 45–46
 interpretation, 61
 stepped chirp waveform processing, 43–45, 45*f*
- Frequency modulation (FM), 739
- Frequency shift keying (FSK), 70
- FSA. *See* Forward-scatter alignment (FSA)
- FSK. *See* Frequency shift keying (FSK)
- G**
- Gait velocitygram, 710
- GAL. *See* Gradient adaptive lattice (GAL)
- Galileo transmitters, 740
- GAMP. *See* Generalized AMP (GAMP)
- Gating, 673–674
- Gaussian distribution, 158, 159*f*
- Gaussian *g-snake* model, 712
- Gaussian noise, disadvantage of, 541
- Gaussian pulse, 194–195, 195*f*
- General Atomics Aeronautical Systems, 707
- Generalized AMP (GAMP), 180
 matrix uncertainty and, 182
- Generalized likelihood ratio test (GLRT), 188, 649
- Generalized sidelobe canceller (GSC), 482–483
 adaptive jammer cancellation, 427, 428*f*
 block diagram, 482*f*
- Geocoding, InSAR, 375
- Geometric theory of diffraction (GTD), 186
- GeoSAR, 385
- GLObal NAVigation Satellite System (GLONASS), 740
- Global System for Mobile Communications (GSM), 740
- GLONASS. *See* GLObal NAVigation Satellite System (GLONASS)
- GLRT. *See* Generalized likelihood ratio test (GLRT)
- GMTI. *See* Ground moving target indication (GMTI)
- “Golden Rule,” 512
- Goldstein-Zebker-Werner (GZW) algorithm, 368–370
- GPR imaging. *See* Ground-penetrating radar (GPR) imaging
- GPU. *See* Graphical processing units (GPU)
- Gradient adaptive lattice (GAL), 757, 758
- Graphical processing units (GPU), 409, 410
- Grating lobes
 effects, 517–520
 sub-arrays and, 518–520, 518*f*, 519*f*
- Grazing angle, selection of, 221
- Greedy algorithms, 176–178, 178*f*
- Green’s function, 153
- Ground clutter, 706–707
 in radar polarimetry, 617
 space-time properties of, 472–474
 covariance matrix, 472
 Doppler frequencies, 472–473
 forward-looking array (FLA), 473
 Hadamard identity, 473
 MCRAM, 473–474
 sidelooking array (SLA), 473
- Ground flash, 232
- Ground moving target indication (GMTI), 2
 clutter, 96
 MIMO, 140–141
 radars, 707, 722
 slow-moving target detection and, 722, 722*f*

SINR loss for, 141, 142*f*
 Ground-penetrating radar (GPR) imaging, 183, 188–191, 190*f*, 191*f*
 Ground plane. *See* Slant/ground plane operations, stripmap SAR
 Ground surveillance radars, performance specifications, 706–707, 706*t*
 Group delay, 53
 instantaneous frequency and, 50, 51*f*, 52
 GSC. *See* Generalized sidelobe canceller (GSC)
 GSM. *See* Global System for Mobile Communications (GSM)
 GTD. *See* Geometric theory of diffraction (GTD)
 Guard band channels, 571
 Guard channel, SLB, 556
 Guard gates, 575–576, 575*f*
 GZW. *See* Goldstein-Zebker-Werner (GZW) algorithm

H

Hamming taper, 51
 Hard thresholding operation, 175
 Harmonic analysis, 54
 Hausdorff distance, 648–649
 Helmholtz equation, 153
 Herglotz kernel method, 191–193
 Hermitian matrix, 471–472
 Heterodyne receiver, 28
 Heterogeneous clutter, 492
 Hidden Markov model (HMM)
 SAR, 650–651
 High-definition vector imaging, 643
 High-performance embedded computing (HPEC), 87
 High range resolution (HRR), 2, 4, 34–36, 60*f*, 631, 658
 image formation, ISAR, 653
 SF waveform, 59
 High resolution elevation (HRE), 344
 High-resolution range profiles (HRRPs)
 observing feature set, 659–660
 selecting feature set, 659
 testing feature set, 660–661
 High-resolution terrain information (HRTI), 344
 High time bandwidth waveforms, periodic vs., 517, 517*f*
 Hilbert-Schmidt distance, 642
 HMM. *See* Hidden Markov model (HMM)
 Hold phase, in pull-off sequence, 544
 Home-on-jam (HOJ), 541, 580
 HOMHT. *See* Hypothesis-oriented MHT (HOMHT)
 Homodyne (single-stage) receiver, 28
 Hough transform, 649
 ISAR, 655
 HPEC. *See* High-performance embedded computing (HPEC)
 HRR. *See* High range resolution (HRR)
 HRRPs. *See* High-resolution range profiles (HRRPs)
 HRTI. *See* High-resolution terrain information (HRTI)
 Human-dismount detection, 705–729
 acronyms, 710
 characterization, human returns, 710–719
 body coordinate system, defined, 713–714, 713*f*
 Boulic-Thalmann model, 713–718, 713*f*, 715*f*, 716*f*, 717*t*
 expected radar return from human target, 711–712, 712*f*
 heel strike (HS), 713
 human kinematic models, 712–716, 713*f*, 715*f*–716*f*
 lifting of toe (TL), 713
 RCS modeling, 718–719, 719*t*
 relative time parameter, 714
 ground clutter, 706–707
 key points, 708
 knowledge exploitation for detection and classification, 727–729, 727*f*
 notation, 708–709
 overview, 705–708
 spectrogram analysis, human returns, 719–721, 721*f*
 technical challenges, 722–727
 inherent output SNR losses, 725–727, 725*f*, 727*f*
 mitigate clutter effects with STAP, 722–724, 724*f*
 slow-moving target detection and GMTI radars, 722, 722*f*
 Human kinematic models, 712–716, 713*f*, 715*f*–716*f*
 Human micro-Doppler signature, 710–711
 Huynen fork, 606
 Hypothesis-oriented MHT (HOMHT), 682–685, 682*f*, 682*t*, 683*t*, 684*t*

I

IBW. *See* Instantaneous bandwidth (IBW)
 ICA. *See* Independent component analysis (ICA)
 ICX Technologies, 706
 IDFT. *See* Inverse DFT (IDFT)
 IDR. *See* Instantaneous dynamic range (IDR)
 IEDs. *See* Improvised explosive devices (IEDs)
 IF. *See* Intermediate frequency (IF)
 IFM. *See* Instantaneous frequency measurement (IFM) receiver
 IHT. *See* Iterative Hard Thresholding (IHT)
 Image frequency, for mixer, 568
 Image jamming techniques, 544
 Image pair generation and registration, 363–365, 364*f*
 Image rejection, receiver-based EP and, 568
 Imaging, SAR
 examples, 251*f*
 metrics, 240–244
 phase error effects on. *See* Phase error effects
 polar formatting algorithm, 235–239, 236*f*, 237–238*f*, 239*f*
 quality of, 244
 reconstruction, 234–240
 tomographic paradigm, 234–235, 235*f*
 Imaging properties, 264–271
 DBS, 269–270, 271*f*
 image formation survey, 268*t*, 269
 PSR, 266–267, 267*f*, 269
 resolution and sampling, 265–266, 265*f*
 Imaging radar mode, 2, 4–5. *See also* High range resolution (HRR); Synthetic aperture radar (SAR)
 IMM. *See* Interacting multiple model (IMM)
 Improvised explosive devices (IEDs), 707
 Independent component analysis (ICA), 646
 Inertial navigation system (INS), 250
 In-phase/quadrature (I/Q) samples, 404
 Input-output (I/O) throughput, 409, 410*f*
 INS. *See* Inertial navigation system (INS)
 InSAR. *See* Interferometric synthetic aperture radar (InSAR)
 InSARE. *See* InSAR elevation (InSARE) systems
 InSAR elevation (InSARE) systems, 385
 Instantaneous AGC (IAGC), 569
 Instantaneous bandwidth (IBW), 539
 Instantaneous dynamic range (IDR), wide, 569
 Instantaneous frequency
 and group delay, 50, 51*f*, 52
 for LFM waveform, 48, 49*f*
 model for, 54, 55*f*
 for NLFM waveform, 48, 49*f*
 slope of, 51
 Instantaneous frequency measurement (IFM) receiver, 537–538
 Integrated sidelobe energy (SLE), 250
 Integrated sidelobe ratio (ISR), 75, 243–244, 244*t*, 250
 lower, MMFs and, 75–76
 Integration angle, 225
 Interacting multiple model (IMM), 13
 estimators, 688–690, 689*f*

- Interference
 - colored noise, 92–95, 93*f*, 94*f*
 - spectrum of, 94*f*
 - covariance matrix, 110
 - estimation, 110*f*
 - Interference-to-noise ratio (INR), 471
 - Interferogram phase statistics, InSAR, 379–380, 380*f*
 - Interferometric synthetic aperture radar (InSAR), 10
 - airborne systems, 385–386
 - coherent exploitation, 386–392
 - ATI, 389–392
 - CCD, 388–389, 389*f*
 - terrain motion mapping, 386–388, 387*f*–388*f*
 - DTM. *See* digital terrain model (DTM)
 - elevation measurement, 344–359. *See also* Elevation measurement, InSAR
 - error sources, 375–382
 - coherence and statistical, 376–381
 - systematic, 381–382
 - images, example, 338*f*
 - operational considerations, 359–362
 - one-pass vs. repeat-pass, 359–360, 360*f*
 - spaceborne vs. airborne, 360–361
 - overview, 337–338
 - processing steps, 362–375
 - absolute elevation estimation, 374–375
 - geocoding, 375
 - image pair generation and registration, 363–365, 364*f*
 - multibaseline, 373
 - orthorectification, 375
 - terrain map estimation, 374
 - two-dimensional phase unwrapping, 366–373
 - unwrapped phase estimation, 374
 - wrapped interferometric phase difference, 365–366
 - spaceborne systems, 384–385
 - Interleaved scanning, 515–516, 515*f*, 516*f*
 - bistatic search and track, 516, 516*f*
 - STAP, 516, 516*f*
 - surveillance, 516, 516*f*
 - Intermap Technologies, Ltd., 385
 - Intermediate frequency (IF), 26
 - high first, 568
 - Intermodulation products, 569
 - Interpulse modulation, 538
 - Intrapulse modulation, 538
 - Intra-pulse scanning (fast scanning), 505–506, 506*f*
 - Invariant histograms, SAR, 649
 - Inverse DFT (IDFT), 749
 - Inverse gain (IG), 544
 - CONSCAN process and, 559, 559*f*
 - Inverse synthetic aperture radar (ISAR), 2, 631
 - candidate features, 654–656
 - adaptive contourlet transform, 655
 - edge detection and watershed segmentation, 655
 - Hough transform, 655
 - log-spiral transform, 655–656
 - ships, features, 656
 - Wavelet transform, 655
 - image formation, 652–653
 - angular velocity and acceleration, estimation, 652–653
 - motion compensation, 652
 - time frequency processing, 653
 - using different frequency subspaces, 653
 - using high range resolution profiles, 653
 - scattering models, 653–654
 - Euler decomposition, 654
 - filtering and phase analysis, 653
 - Pol-CLEAN technique, 654
 - 3-D scatterer models, 654
 - IRLS. *See* Iterative reweighted least squares (IRLS)
 - Irregular subarray, 412–413, 413*f*
 - digitally restereed antenna pattern, 417, 417*f*
 - ISAR. *See* Inverse synthetic aperture radar (ISAR)
 - Isotropic scatterers model, 509n6
 - ISR. *See* Integrated sidelobe ratio (ISR)
 - ISTA. *See* Iterative shrinkage-thresholding algorithm (ISTA)
 - Iterative reweighted least squares (IRLS), 176
 - Iterative reweighting schemes, 175–176
 - Iterative shrinkage-thresholding algorithm (ISTA), 173
 - Itrative Hard Thresholding (IHT), 175
- J**
- JAFF. *See* Jammed chaff (JAFF)
 - Jammed chaff (JAFF), 544
 - Jammer cancellation, adaptive algorithms
 - for, 419–435
 - constrained optimization, 425–429
 - derivative constraints, 428
 - eigenvector constraints, 428–429
 - GSC, 427–428, 428*f*
 - LCMV, 426–427, 427*f*
 - linear array model, 419, 419*f*
 - maximum SINR, 423–425
 - noise vector, 420
 - performance metrics, 432–435
 - antenna patterns, 434
 - cancellation ratio (CR), 433
 - cumulative distribution functions, 434–435, 435*f*
 - residue, 433
 - SINR, 433–434
 - steering vector, 420
 - vector notation, 420
 - weight estimation, 429–432
 - Cholesky factorization, 432
 - key considerations, 429–430
 - sample matrix inversion, 430–431
 - SINR loss, 430, 431*f*
 - weight jitter stabilization, 431–432, 431*f*
 - weight vector, 420–421, 421*f*
 - Wiener filter, 422–423
- Jammers, 534. *See also* Specific entries
 - received power
 - computation, 552
 - example, 553, 553*f*
 - T/R isolation, 537, 555
 - types, 535–540
 - coherent, 536*f*, 537–540, 539*f*
 - noncoherent, 536–537, 536*f*
- Jamming, additive noise, 110
- Jam strobe, 580
- Jam-to-noise ratio (JNR)
 - for noise jammer, 547
- Jam-to-signal ratio (JSR), 544
 - for noise jammer, 547, 548
- JNR. *See* Jam-to-noise ratio (JNR)
- Jog detector, 559
- Joint domain localized (JDL) algorithm, 486–487, 486*f*
- Joint spatial-frequency domain wideband beamformer, 444–446, 444*f*
- Jones matrix, 602
- Jones vector, 598
- Jonker-Volgenant-Castanon (JVC) algorithm, 675
- JSR. *See* Jam-to-signal ratio (JSR)
- JVC. *See* Jonker-Volgenant-Castanon (JVC) algorithm
- K**
- Ka-band radar, images of, 222, 224*f*
 - Kalman filter, 675
 - Karhunen-Löve transform (KLT), 470
 - Kennaugh matrix, 602, 608–609
 - KLT. *See* Karhunen-Löve transform (KLT)
 - Kruskal rank, 162–163
 - Kullback-Leibler distance, 635
- L**
- Lagrange multiplier theory, 168
 - Lagrangian relaxation, 685
 - LaMP (lattice matching pursuit), 181

- Layover and foreshortening, InSAR, 355–358, 356f–357f, 358f
- LCMV. *See* Linear constrained minimum variance (LCMV)
- Leading-edge track (LET), 576, 576f
- Least jammed frequency selection, 571–572
- Least squares (LS)
 method, 369–371, 370f, 371f
 solution, 76
- LET. *See* Leading-edge track (LET)
- LFM. *See* Linear frequency modulation (LFM)
- Likelihood-based classifier, SAR, 649
- Likelihood ratio test (LRT), 3
- Linear coherent repeater, 582
- Linear constrained minimum variance (LCMV), 426–427
- Linear constraints, optimum MIMO waveform, 104–105, 105f
- Linear frequency modulation (LFM), 273, 446–449, 447f, 448f, 565
 waveform, 19–20, 56–57
 chirps, 136–137f, 136–138, 226, 227
 instantaneous frequency for, 48, 49f
 stretch processing and. *See* Stretch processing
 transmission of, 27–28
 unoptimized, 107f
- Linear model
 compressed sensing, 150–151
 regularization, sparsity and, 158–159
- Linear–nonlinear integration, of multiple frequency PBR channels, 791–803, 793f, 795f, 797f–802f, 798t
- Linear phase errors, 246
- Linear polarization, 595, 595f
 Poincaré sphere and, 599–600, 600f
- Linear repeater, 549
- Linear restricted scattering matrix (LSM), 619
- Linear sampling method (LSM), 184, 192
- Linear time-invariant (LTI), 89
- Link equation approach, 545
- Lipschitz constant, 173
- L0 norm, linear model regularization, 152
- L1 norm, linear model regularization, 159–161
- Lobe-on-receive-only (LORO), 559
- Local oscillator (LO), 539
 signals, 408
- Log-spiral transform, ISAR, 655–656
- Long-pulse waveforms, 562–563
- Long Term Evolution (LTE), 740
- Look-through periods, 537
- LORO. *See* Lobe-on-receive-only (LORO)
- Loss in processing gain (LPG), 75
- Low antenna sidelobes, 555
- Low cross polarization antenna, 560–561
- Lower limits, PRF, 318–321
- Low-noise amplifier (LNA), 404
- Low-pass filter, bandwidth, 34
- LPG. *See* Loss in processing gain (LPG)
- LRT. *See* Likelihood ratio test (LRT)
- LS. *See* Least squares (LS)
- LSM. *See* Linear restricted scattering matrix (LSM); Linear sampling method (LSM)
- LTE. *See* Long Term Evolution (LTE)
- LTI. *See* Linear time-invariant (LTI)
- Lynx Advanced Multi-Channel Radar (AMR), 707
- M**
- MACH. *See* Maximum average correlation height (MACH)
- Mahalanobis distance, 674
- Main lobe cancellation (MLC), 557–558
- Main lobe clutter, by moving platform, 501n1
- Majorization minimization (MM) algorithm and FOCUSS algorithm, 176
 ISTA as, 173
- MANs. *See* Metropolitan area networks (MANs)
- Map drift (MD), 253
- Masking EA techniques, 533, 534, 540–542
 coherent, 542
 noncoherent, 540–541
- Matched filter/filtering, 44, 45, 286–288, 725
 AWGN, 287
 convolution integral, 288
 one-dimensional impulse response, 286, 286f
 response
 for CW pulse, 64, 64f
 quadrature waveforms, 74
 SNR, 287
 whitening, 90, 91
- Matching Pursuits (MP) algorithm, 177
- MATLAB, 468
- Maximum average correlation height (MACH)
 SAR, 651
- MCARAM. *See* Multi-Channel Airborne Radar Measurements (MCARM) Program
- MD. *See* Map drift (MD)
- MDV. *See* Minimum detectable velocity (MDV)
- Mean square error (MSE), 179
 SAR, 648
- Measurement-level fusion, 691, 692
 challenges, 692, 694
- Measurement-to-track association, 669, 672–675
- MESAR 2, 410, 411, 411f
- Metropolitan area networks (MANs), 781
- MHT. *See* Multiple hypothesis tracking (MHT)
- Microwave power module (MPM)
 amplifier, 540
- Migrating target indicator (MiTI), 523–524, 523f
- Millimeter wave (mmW) SAR, 640
- MIMO radar. *See* Multi-input, multi-output (MIMO) radar
- Minimum detectable velocity (MDV), 8, 141, 707, 722
- Minimum mean square error (MMSE), 178, 179, 482
- Minimum shift keying (MSK), 70
- Minimum variance beamformer, 481–482
- Minimum variance distortionless response (MVDR), 470, 470f
- Mismatched filters (MMFs), 21, 52, 75–81
 Doppler shift and, 78, 79f
 lowest ISR, 75–76
 MISR filters, 76–78, 77f, 77t, 78f, 78t
 performance metrics, 75
 performance of, 78t
 phase codes used in, 77t
 to reduce PSR, 79–80, 80f
 and sidelobe response, 78–79
 variables, 25
- MiTI. *See* Migrating target indicator (MiTI)
- MLC. *See* Main lobe cancellation (MLC)
- MM algorithm. *See* Majorization minimization (MM) algorithm
- MMFs. *See* Mismatched filters (MMFs)
- MMSE. *See* Minimum mean square error (MMSE)
- MNR. *See* Multiplicative noise ratio (MNR)
- Modulated straight-through repeater, 540
- Monopulse angle measurement, 559–560, 560f
- Monopulse angle tracking radar, 544
- Monopulse ratio, of difference and sum beams, 557
- Morozov's discrepancy principle, 190
- Motion compensation, 652
- Motion-through-resolution-cells (MTRC), 9
- Moving and stationary target acquisition and recognition (MSTAR), 184, 638
- Moving target imaging, 185–187, 185f, 186f
- Moving target indication (MTI), 453, 560.
 See also Space-time adaptive processing (STAP)
 beamforming, 459–464
 Doppler processing, 464–466
 radar, 2–4, 135

- Moving target indication (MTI) (*cont.*)
 ambiguity function, 153–155
 data collection scenario, 156*f*
 spatial sampling, 459–464
 technique, 756
 temporal sampling, 464–466
- MP algorithm. *See* Matching Pursuits (MP) algorithm
- MPM. *See* Microwave power module (MPM) amplifier
- MQ-1C SkyWarrior/Gray Eagle unmanned aircraft system, 707
- MRA. *See* Multi-resolution analysis (MRA)
- MSE. *See* Mean square error (MSE)
- MSK. *See* Minimum shift keying (MSK)
- MSTAR. *See* Moving and stationary target acquisition and recognition (MSTAR)
- MTI. *See* Moving target indication (MTI)
- MTRC. *See* Motion-through-resolution-cells (MTRC)
- Mueller matrix, 607–611
- Multi-Channel Airborne Radar Measurements (MCARM) Program, 473
- Multichannel phased array radar, 155–157, 156*f*
- Multichannel processing
 for PBR detection performance improvement, 791–813
 adaptive antenna array, 808–813, 810*f*–812*f*, 814*f*
 linear–nonlinear integration, 791–803, 793*f*, 795*f*, 797*f*–802*f*, 798*t*
 multiple polarimetric passive radar channels exploitation, 803–808, 805*t*, 806*f*–809*f*
- Multi-input, multi-output (MIMO) radar, 1, 8, 119–142
 adaptive, 109–113
 angle ambiguity function for, 141
 array factor of, 131, 132*f*
 beampatterns, 142
 characteristics of, 120
 cohere-on-target, 112–113, 113*f*
 configuration, 6
 distributed aperture, 141
 dynamic calibration, 111–112, 112*f*
 GMTI, 140–141
 notation, 120–121
 optimum waveform design
 for ACN, 89–95, 89*f*, 90*f*, 93*f*, 94*f*
 linear constraints, 104–105, 105*f*
 for SCR, 95–99, 96*f*, 97*f*
 for target identification, 99–104, 100*f*, 101–103*f*
 organization, 119–120
 overview, 119, 121–122, 121*f*
 range response, 136
 SAR system, 138–140, 140*t*
 signal correlation matrix, 127–129, 129*f*
 signal model, 124–127, 126*f*
 spatial beamforming, 130
 virtual array, 122–124, 124*f*
 waveforms for, 135–138, 136–137*f*
 with widely separated antennas, 122
- Multipath, interference
 colored noise, 92–95, 93*f*, 94*f*
- Multiple hypothesis tracking (MHT), 13, 678, 682–686
 HOMHT, 682–685, 682*f*, 682*t*, 683*t*, 684*t*
 TOMHT, 685–686
- Multiple polarimetric passive radar channels exploitation, PBR system, 803–808, 805*t*, 806*f*–809*f*
- Multiple-pulse waveform, 563
- Multiple simultaneous frequency radiation, 566–567
- Multiple spot noise, 541
- Multiplicative noise, 242–244
 ambiguity-to-signal ratio, 242–243
 integrated sidelobe ratio, 243–244, 244*t*
 quantization noise ratio, 244
- Multiplicative noise ratio (MNR), 9, 241
- Multi-resolution analysis (MRA), 645–646
- Multisensor tracking, 691–694
 measurement-level fusion, 691, 692
 challenges, 692, 694
 sensor fusion architectures, 691–692
 track-level fusion, 692, 693*f*
 challenges, 692–693, 694
- Multistatic radar, topology, 6
- Multitargets, identification of, 103–104, 103*f*
- Multitarget tracking, 677–690
 measurement-to-track data association, 677–686
 cluster tracking, 686
 feature-assisted tracking, 678–682, 681*f*
 multiple hypothesis tracking, 682–686, 682*f*, 683*t*, 684*t*. *See also* Multiple hypothesis tracking (MHT)
 modeling target dynamics, 686–690, 687*f*
 interacting multiple model (IMM) estimators, 688–690, 689*f*
 VS-IMM estimators, 690
- Mutual coherence, 165–166
- Mutual-protection technique, 545
- MVDR. *See* Minimum variance distortionless response (MVDR)
- Myriad waveforms, 226
- ## N
- Nadir return, 232
 constraints on, 233
- Narrow antenna beamwidth, 558
- Narrowband Doppler noise (NBDN), 542
- Narrowband repeater noise (NBRN), 542
- Narrow-pulse range resolution, 564
- National Image Interpretability Rating Scale
 for radar imagery (RNIIRS), 244, 245*t*
- NBDN. *See* Narrowband Doppler noise (NBDN)
- NBRN. *See* Narrowband repeater noise (NBRN)
- NCP. *See* Noise cover pulse (NCP)
- Nearest neighbor, SAR, 650
- NESTA. *See* Nesterov's Algorithm (NESTA)
- Nesterov's Algorithm (NESTA), 167
- NESZ. *See* Noise-equivalent sigma zero (NESZ)
- Netflix problem, 182
- Network flow method, 371–372
- Neural nets, SAR, 649–650
- Neyman-Pearson framework
 ATR, performance prediction, 635–636
- NLFM waveforms. *See* Nonlinear frequency modulated (NLFM) waveforms
- NLMS algorithm. *See* Normalized least mean squares (NLMS) algorithm
- ## Noise
- additive, 242
- jammers, 536–537
 computation example, 548–549, 548*t*, 549*f*
 formulas, 547–548
- multiplicative, 242–244
- quantization, 244
- thermal, 242
- thermal and quantization, 376–377
- Noise cover pulse (NCP), 541, 571
- Noise-equivalent sigma zero (NESZ), 242
- Noncoherent detection channel, 571
- Non-coherent integration
 coherent vs., 512, 513*f*
- Noncoherent jammer, 12, 536–537, 536*f*
- Noncoherent masking techniques, 540–541
- Noncooperative bistatic radar, 6
- Nonlinear constraints, optimum MIMO waveform, 106–109, 106*f*, 107*f*, 109*f*
- Nonlinear frequency modulated (NLFM) waveforms, 7, 20, 21, 48–58
 coefficients used in generation of, 55, 55*t*
 and constant modulus, 108–109, 109*f*
 creation, example of, 52–57, 54*f*, 55*f*, 55*t*, 56*f*

- design approaches, 51–52
- Doppler analysis, 57, 57*f*, 58*f*
- functional relationships, 49–51
- instantaneous frequency for, 48, 49*f*
- inverse functions, 50
- parametric equations, 50–51
- principle of stationary phase, 49–50
- spectrum, 56, 56*f*
- variables, 23–24
- Nonstationary clutter, 492
- Nonsymmetric noise target, 610
- Nonuniform FFT (NUFFT), 186
- Norden Systems, Inc., 385
- Normalized angle, 97
- Normalized least mean squares (NLMS)
 - algorithm, 757–758
- Northrop Grumman, 707
- Notch filtering, receiver-based EP and, 569–570
- N_{sc} LFM pulses, 40
- NUFFT. *See* Nonuniform FFT (NUFFT)
- Null hypothesis, radar detection, 2–3
- Null polarizations, 605, 606
- Nyquist array, 124
- Nyquist sampling rate, 150
- O**
- OFDM modulation. *See* Orthogonal frequency-division multiplexing (OFDM) modulation
- Off-board category, angle track deception and, 544
- OMP algorithm. *See* Orthogonal Matching Pursuits (OMP) algorithm
- On-board category, angle track deception and, 544
- One-pass vs. repeat-pass operations, 359–360, 360*f*
- Optimum pulse shape, for SCR
 - maximization, 98–99
- Optimum waveform, MIMO
 - for ACN, 89–95, 89*f*, 90*f*, 93*f*, 94*f*
 - linear constraints, 104–105, 105*f*
 - nonlinear constraints, 106–109, 106*f*, 107*f*, 109*f*
 - for SCR, 95–99, 96*f*, 97*f*
 - for target identification, 99–104, 100*f*, 101–103*f*
- Oracle solution, 177
- Ordered statistic CFAR (OS-CFAR), 573–574
- Orthogonal frequency-division multiplexing (OFDM) modulation, 739, 780, 781
- Orthogonal Matching Pursuits (OMP)
 - algorithm, 177–178, 178*f*
- Orthogonal waveforms
 - advantage of, 133
 - phased array vs., 131–134, 132*f*, 132*t*, 133–134*f*
- Orthorectification, InSAR, 375
- OS-CFAR. *See* Ordered statistic CFAR (OS-CFAR)
- Oscillators
 - over sweep, 33–34, 34*f*
 - receive, 28, 29*f*
- OTHR. *See* Over-the-horizon radar (OTHR)
- Overlapped subarray, 413, 413*f*, 414, 417–418, 417*f*
 - digitally resteeered antenna pattern, 417–418, 417*f*–418*f*
- Over-the-horizon radar (OTHR), 6
- P**
- Pareto frontier, 169, 170*f*
- Passive bistatic radar (PBR), 14, 739–813
 - acronyms, 746–747
 - ambiguity function, 742–743
 - 2D-CCF, evaluation of. *See* Two-dimensional cross-correlation function (2D-CCF), in PBR
 - direct signal and multipath/clutter cancellation techniques, 755–766
 - ECA-B approach, 760–761, 760*f*
 - ECA-B&S algorithm, 761–766, 762*f*, 763*t*, 764*f*–765*f*
 - extensive cancellation algorithm, 758–760, 759*f*
 - key points, 744
 - multichannel processing for detection
 - performance improvement, 791–813
 - adaptive antenna array, 808–813, 810*f*–812*f*, 814*f*
 - linear–nonlinear integration, 791–803, 793*f*, 795*f*, 797*f*–802*f*, 798*t*
 - multiple polarimetric passive radar channels exploitation, 803–808, 805*t*, 806*f*–809*f*
- notation, 744–746
- overview, 740–744
- processing scheme
 - advanced, 741–742, 742*f*
 - basic, 740–741, 741*f*
- reference channel, 740
- signal processing techniques, for
 - reference signal cleaning and reconstruction, 766–775
 - constant modulus algorithm using analog modulation, 767–771, 768*f*, 770*f*, 771*t*, 772*f*
 - digital transmissions, 771–775, 774*f*
 - surveillance channel, 740
 - transmitters, 739–740
- Passive conical scanning, 558–559, 559*f*
- Passive radar ATR, 656–658
 - identifying target set, 657
 - observing feature set, 657–658
 - selecting feature set, 657
 - testing feature set, 658
- Path-following approach, 366–369, 368*f*, 369*f*
- PBR. *See* Passive bistatic radar (PBR)
- PC. *See* Pulse compression (PC)
- PCA. *See* Principal component analysis (PCA)
- PCI. *See* Principal components inverse (PCI)
- PDA. *See* Phase difference autofocus (PDA)
- PDF. *See* Probability density function (PDF)
- Peak sidelobe ratio (PSR), 53, 75
 - MMFs and, 79–80, 80*f*
- Peak-to-noise floor ratio (PNFR), 762–763
- Peak-to-sidelobe ratio (PSLR), 762–763
- Peak transmit power, 125
- Pearson distance, 649
- Pedestrian detection. *See* Human–dismount detection
- PEMS. *See* Predict, extract, match, and search (PEMS) modules
- Penalized least squares problem
 - basis pursuit de-noising, 168–170, 170*f*
 - Pareto frontier, 169, 170*f*
- Penetration applications, 329–330, 330*f*
- Periodic bandwidth waveforms
 - high time vs., 517, 517*f*
- Periodic signal, 54
- Perishability, EP, 554
- Persistent scatterers, SAR, 647–648
- PFA. *See* Polar formatting algorithm (PFA)
- PGA. *See* Phase gradient autofocus (PGA)
- Phase analysis, ISAR, 653
- Phase-coded waveform, 20
- Phased array antenna, 404
- Phased array radar, 119
 - elements transmission in, 123
 - gain of, 131
 - multichannel, 155–157, 156*f*
 - vs. orthogonal waveforms, 131–134, 132*f*, 132*t*, 133–134*f*
- Phase difference autofocus (PDA), 252–253
- Phase error effects, 244–250
 - linear, 246
 - overview, 244–246
 - quadratic, 246–247, 247*f*
 - sinusoidal, 247–248, 248*f*
 - wideband, 248–250, 249*f*
- Phase gradient autofocus (PGA), 253, 254*f*
- Phase modulation (PM), 538
- Phase rotation, SF waveform and, 59, 61
- Plane. *See* Slant/ground plane operations, stripmap SAR
- PNFR. *See* Peak-to-noise floor ratio (PNFR)

POI. *See* Probability of intercept (POI)
Poincaré sphere, 598–601, 599*f*
 circular polarization, 600, 600*f*
 defined, 599
 elliptical polarization, 600–601, 600*f*
 linear polarization, 599–600, 600*f*
Point spread function (PSF), 134, 134*f*, 153
Point spread response (PSR), 261, 266–267, 267*f*, 269
 DBS and, 269
 linear phase error and, 246*f*
 matched filtering, 315
 quadratic phase error and, 247*f*
 range migration, 278–279, 279*f*
RDA. *See* Range-Doppler algorithms (RDA)
RSA and, 289, 290–291, 290*f*
slant/ground planes, 323–324
squinted stripmap, 326
Stolt interpolation, 312
two-dimensional function, 286, 286*f*
wideband phase error, 248–250, 249*f*
Polar formatting algorithm (PFA), 9, 235–239, 236*f*, 237–238*f*, 239*f*
 functionality of, 234
 implementation of, 238
Polarimetric whitening filter (PWF), 643–644
Polarimetry, radar. *See* Radar polarimetry
Polarization, 594–601. *See also* Specific entries
 defined, 594
 fork, 606
 Poincaré sphere. *See* Poincaré sphere
 scattering matrix and
 optimal, 604–607
 partial, 607–611
 states, defined, 598
 stokes parameters, 598–601
 wave. *See* Wave polarization
Polarization scattering matrix (PSM), 12–13
Polar-to-rectangular interpolation, 236–237
Pol-CLEAN technique, ISAR, 654
Post-detection (noncoherent)
 integration, 563
Post-Doppler STAP, 487–488, 487*f*
Power spectral density (PSD), 786
PPIN. *See* Pulse-present indicator (PPIN)
Precipitation clutter, 616
Predict, extract, match, and search (PEMS) modules, 638
Pre-Doppler STAP, 488–489, 488*f*
PRF. *See* Pulse repetition frequency (PRF)
PRI. *See* Pulse repetition interval (PRI)
Principal component analysis (PCA), 646
Principal components inverse (PCI), 489
Principal eigenvalue approach, 812–813

Principle of stationary phase (PSP), 49–50
Principles of Modern Radar: Basic Principles, 631
Probability density function (PDF), 3, 100
Probability of intercept (POI), 562
Projection slice theorem, 235
PSD. *See* Power spectral density (PSD)
PSF. *See* Point spread function (PSF)
PSLR. *See* Peak-to-sidelobe ratio (PSLR)
PSM. *See* Polarization scattering matrix (PSM)
PSP. *See* Principle of stationary phase (PSP)
PSR. *See* Peak sidelobe ratio (PSR); Point spread response (PSR)
Pull-off, defined, 544
Pulse
 CW, 59
 frequency shift, 43
 phase corrected, 43
 short, transmission of, 99
Pulse compression (PC), 545
 gain, stretch processing and, 37–38
 range resolution, 564–565
 waveform, 19–80
 receivers in, 28–30
 stretch processing. *See* Stretch processing
Pulse-deinterleaving algorithm, 538
Pulse-descriptor word (PDW), 538
Pulse-present indicator (PPIN), 537
Pulse repetition frequency (PRF), 40, 226
 lower limits, 318–321
 timing constraints, 231–233, 233*f*
 upper limits, 321–322
 values for SIR-B radar, 233*f*
 variation, 567
Pulse repetition interval (PRI), 7, 40, 231, 464, 538, 711
 agility, 564
 threat parameters, 534
 tracker, 542
Pulse-to-pulse frequency agility, 564, 566
Pulse-to-pulse jitter, 567
Pulse-to-pulse stagger, 567
Pulse width (PW) measurement, 537
PWF. *See* Polarimetric whitening filter (PWF)

Q

QNR. *See* Quantization noise ratio (QNR)
QPE. *See* Quadratic phase error (QPE)
Quadratic phase error (QPE), 246–247, 247*f*
Quadrature codes, 20, 21
 13-element Barker code conversion to, 71*t*
 variables, 24–25
Quadrature waveforms, 7, 70–75
 ambiguity surface of, 74

 match filtered response, 74
 spectrum, 72–74, 73*f*, 74*f*
Quantization noise, 376–377
Quantization noise ratio (QNR), 244

R

Radar à Impulsion et Antenne Synthétiques (RIAS) concept, 514, 514*f*
Radar cross section (RCS), 4, 13–14
 modeling, of human in motion, 718–719, 719*t*
 statistics, 577
 step detection, 578
Radar polarimetry, 589–622
 acronyms, 593
 applications of, 611–618
 ground clutter, 617
 rain clutter, 616–619
 sea clutter, 617–618
 targets. *See* Targets, in radar polarimetry
 foundations of, 590
 key points, 591–592
 list of symbols, 592–593
 organization, 591
 overview, 589–593
 polarization, 594–601
 defined, 594
 Poincaré sphere. *See* Poincaré sphere
 stokes parameters, 598–601
 wave. *See* Wave polarization
 scattering matrix in. *See* Scattering matrix
Radar(s)
 bistatic. *See* Bistatic radar
 concepts, 1–2
 detection, 2, 3–4
 modes, 2–5. *See also* Imaging radar mode; Moving target indication (MTI)
 multi-input, multi-output. *See* Multi-input, multi-output (MIMO)
 radar
 multistatic, 6
 peak transmit power, 125
 post-processing, 2, 12–13
 range gates in, 64
 sensors vs., 705
 topologies, 5–6
RADARSAT-1, 329
Radar system losses, 546–547
Radar warning receiver (RWR), 534
Radial velocity, estimates of, 578
Radio frequency interference (RFI), 2, 10–11
Radio frequency (RF)
 band designations, 535*t*
 bandwidth, 226
 combiner networks, 404

- preselection, 567–568
 - spatio-temporal, 87
 - tomographic system, 152–153, 152*f*
 - transmission frequency, 153
 - Radon transform, 641
 - Rain clutter, in radar polarimetry, 616–617
 - Raj, R. G., 712
 - Random beam scanning, 558
 - Range ambiguities, 232, 243
 - Range bin masking, 542
 - Range-Doppler algorithms (RDA), 306
 - approximations, 299–302
 - depth of focus, 296–298
 - frequency-domain PSR for, 298–299
 - full RDA, 292–296
 - modified, 302–304
 - notes on, 304
 - Range-Doppler coupling, 39
 - Range-Doppler track comparison, 578–579
 - Range false targets (RFT), 542, 543
 - Range gate capture (RGC), 543–544
 - Range gate pull-in (RGPI), 543–544
 - Range gate pull-off (RGPO), 543–544, 575
 - Range gate stealing (RGS), 543–544
 - Range gate walk-off (RGWO), 543–544
 - Range history data, 229
 - Range migration algorithm (RMA), 10, 261, 276, 305–318
 - chirp scaling algorithm (CSA), 315–318
 - coordinate systems and frequencies, 306–307
 - RDA, 306
 - simulated examples, 308–312, 308*f*–310*f*, 311*f*–312*f*
 - Stolt interpolation, 307, 313–315, 314*f*
 - Range migration compensation (RMC), 276–285, 277*f*, 279*f*, 280*f*, 281*f*–283*f*, 284*f*, 285*f*
 - Range rate, estimates of, 578
 - Range resolution, 32–33, 562
 - frequency resolution and, 36
 - narrow-pulse, 564
 - pulse compression, 564–565
 - synthetic aperture radar, 226
 - Range response, MIMO radar, 136
 - Range sidelobe control, 2D-CCF
 - in WiFi-based PBR, 776–780, 777*f*, 778*t*, 780*f*
 - in WiMAX-based PBR, 780–786, 781*t*, 782*f*, 783*f*, 785*f*
 - Range sidelobes, suppression of, 38–39, 38*f*, 63
 - Range stacking algorithm (RSA), 288–292
 - convolution filtering, 290–291
 - FFT-based convolution in, 292
 - Fourier transform (FFT), 290
 - image formation algorithm, 288
 - matched filter image former, 289
 - PSR, 289
 - signal flow diagram, 291*f*
 - Range track deception technique, 543–544
 - Range-walk, 564
 - Rapid Terrain Visualization System (RTV) system, 385–386
 - Rayleigh criterion, 32
 - Rayleigh distribution, 2
 - Rayleigh quotient, 96
 - Rayleigh range resolution, 45
 - SF waveform, 62
 - RCS. *See* Radar cross section (RCS)
 - RDA. *See* Range-Doppler algorithms (RDA)
 - R-dot-v comparison, 578
 - Receiver-based EP, 567–572
 - guard band channels, 571
 - image rejection, 568
 - least jammed frequency selection, 571–572
 - noncoherent detection channel, 571
 - notch filtering, 569–570
 - RF preselection, 567–568
 - wideband limiting, 570–571
 - wide IDR, 569
 - Receiver operating characteristic (ROC) curves
 - ATR and, 634–635
 - Receiver(s)
 - for ACN, 90, 90*f*
 - adaptivity, 109
 - DBF, 407–409
 - deramp, 226–229
 - homodyne, 28
 - joint design, 100*f*
 - optimal, 100*f*
 - processor architecture, 30–31, 31*f*
 - in pulsed system, 28–30
 - sensitivity, 552
 - SF waveform, 60, 60*f*
 - signal and timing relationships, 29*f*
 - Recirculating RF memory loops, 565
 - Rect function, 216, 218
 - Reduced-dimension STAP (RD-STAP), 455, 484–488
 - extended factored algorithm (EFA), 485–486, 485*f*, 487, 488
 - joint domain localized (JDL) algorithm, 486–487, 486*f*
 - Reduced-rank STAP, 489
 - Reference cells, 573
 - Reference profile, InSAR, 355–356
 - Reference signal
 - cleaning and reconstruction, signal processing techniques for, 766–775
 - constant modulus algorithm using analog modulation, 767–771, 768*f*, 770*f*, 771*t*, 772*f*
 - digital transmissions, 771–775, 774*f*
 - Reflectivity function, 154
 - Regular subarray, 412, 413, 413*f*
 - digitally resteered antenna pattern, 416, 416*f*
 - Relaxed projection approach, 106–107, 107*f*
 - Remote sensing, 327–329, 328*t*, 329*f*
 - Repeater, 540. *See also* Specific entries
 - Repeater throughput gain
 - at radar polarization, 551
 - Repeat-pass operations, one-pass vs., 359–360, 360*f*
 - Residual video phase (RVP), 29
 - Residue, 433
 - Resolution, 539
 - Responsive spot noise (RSN), 541
 - Restricted isometry constant (RIC), 163
 - Restricted isometry property (RIP), 163–165
 - belief propagation, 179
 - RF. *See* Radio frequency (RF)
 - RFI. *See* Radio frequency interference (RFI)
 - RFT. *See* Range false targets (RFT)
 - RGC. *See* Range gate capture (RGC)
 - RGPI. *See* Range gate pull-in (RGPI)
 - RGPO. *See* Range gate pull-off (RGPO)
 - RGS. *See* Range gate stealing (RGS)
 - RGWO. *See* Range gate walk-off (RGWO)
 - RIAS. *See* Radar à Impulsion et Antenne Synthétiques (RIAS) concept
 - RIC. *See* Restricted isometry constant (RIC)
 - RIP. *See* Restricted isometry property (RIP)
 - RMA. *See* Range migration algorithm (RMA)
 - RNIIRS. *See* National Image Interpretability Rating Scale for radar imagery (RNIIRS)
 - Robustness, EP, 554
 - ROC. *See* Receiver operating characteristic (ROC) curves
 - Rotman lens, 538
 - RSA. *See* Range stacking algorithm (RSA)
 - RSN. *See* Responsive spot noise (RSN)
 - RVP. *See* Residual video phase (RVP)
 - RWR. *See* Radar warning receiver (RWR)
- S**
- Sample matrix inversion, 430–431
 - Sampling rate, ADC, 26, 37, 41–42, 226–227
 - Sandia National Laboratories, 385, 707
 - SAR. *See* Synthetic aperture radar (SAR)
 - SAS. *See* Space and Aerospace Systems (SAS)
 - Saturated coherent repeater, 582

- Saturated (constant amplitude) video pulse, 537
- Saturated repeater, 549
- SAW. *See* Surface acoustic wave (SAW)
- Scan-on-receive-only (SORO), 559
- Scan with compensation (SWC), 560
- Scatterers, 38
 - on range gate, 64*f*
- Scattering matrix, 590, 601–611
 - bistatic, 602
 - Jones matrix, 602
 - Kennaugh matrix, 602, 608–609
 - measurement of, 618–621, 621*f*
 - Mueller matrix, 607–611
 - polarization
 - optimal, 604–607
 - partial, 607–611
 - Sinclair matrix, 602
 - formulation, 602–604
- Scattering matrix with absolute phase (SMA), 619
- Scattering matrix with relative phase (SMR), 619, 620, 621*f*
- S-CMA. *See* Space-CMA (S-CMA)
- SCR. *See* Signal-to-clutter ratio (SCR)
- Sea clutter, in radar polarimetry, 617–618
- Secondary range compression (SRC), 304
- Sector blanking, 555
- Seduction technique, 543
- Segmentation, 643
- Self-protection jammer (SPJ), 534
- Self-screening jammer (SSJ), 534
- Sensor fusion architectures, 691–692
- Sensor geometry errors, 381
- Sensors, radars vs., 705
- “Separation principle,” 95
- Serrodyning, 538
- SFJ. *See* Stand-forward jammer (SFJ)
- SF waveform. *See* Stepped frequency (SF) waveform
- Shannon-Nyquist sampling theorem, 229
- Shifted delta function, Fourier transform of, 218
- Ships, classification features
 - ISAR imaging, 656
- Short pulse, transmission of, 99
- Short pulse range resolution, 564
- Shuttle Imaging Radar B (SIR-B)
 - PRF values for, 233*f*
- Shuttle Radar Topography Mission (SRTM), 343, 384, 385
- Sidelobe blanker, 435–437, 435*f*–437*f*
- Sidelobe blanking (SLB), 555–556, 556*f*
- Sidelobe cancellation (SLC), 556–557
- Sidelobe canceller, 437–440, 438*f*–440*f*, 813, 814*f*
- Sidelobe false targets (SLFT), 543
- Sidelobes
 - low antenna
 - EP and, 555
 - peak, 56–57, 56*f*
 - range, suppression of, 38–39, 38*f*
 - response, MMFs and, 78–79
 - target suppression, 96–98, 97*f*
- Sidelooking array (SLA), 473
- Signal correlation matrix, MIMO, 126, 127–129, 129*f*, 130
- Signal noise, formulas, 545–546
- Signal processing techniques
 - for reference signal cleaning and reconstruction, 766–775
 - constant modulus algorithm using analog modulation, 767–771, 768*f*, 770*f*, 771*t*, 772*f*
 - digital transmissions, 771–775, 774*f*
- Signal processor-based EP, 572–576
 - alternate CFAR detection, 572–574, 573*f*
 - angle variance, 574
 - data editing, 574–575
 - guard gates, 575–576, 575*f*
 - LET, 576, 576*f*
 - TED, 572
- Signals
 - baseband. *See* Baseband signal
 - digital processing, 7
 - Fourier transform of, 31
 - interpolated, 42
 - periodic, 54
 - sparsity of, 151
 - transmission, MIMO radar, 151
- Signal-to-clutter ratio (SCR)
 - optimum MIMO waveform design and, 95–99, 96*f*, 97*f*
 - optimum pulse shape and, 98–99
- Signal-to-interference-plus-noise ratio (SINR), 7–8, 87, 454, 557, 723, 724
 - gain of the optimum short pulse, 92
 - improvement, 107*f*
 - loss, for GMTI, 141, 142*f*
 - loss for multipath interference scenario, 111*f*
- Signal-to-noise ratio (SNR), 7, 125, 126, 433–434, 748
 - IF factor, 478
 - loss, for human targets in detectors, 725–727, 725*f*, 727*f*
 - loss in, 33
 - maximizing output, 423–425, 424*f*, 478–481, 480*f*, 481*f*
 - MIMO signal correlation matrix, 128
 - output, white noise matched filter and, 91
 - of radar range, 546
 - space-time processing, 453, 454, 476, 477–478
- SII. *See* Stand-in jammer (SIJ)
- SIMO system. *See* Single-input, multiple-output (SIMO) system
- Sinc function, 215–218, 217*f*
- Sinc interpolator, 238–239, 239*f*
- Sinclair matrix, 602
 - formulation, 602–604
- Single-channel transmit waveform, 89
- Single-input, multiple-output (SIMO) system, 124
- Single-input, single-output (SISO) system, 124
- Single-pulse measurements, 538
- Single-pulse target detection, 193–195, 195*f*
- SINR. *See* Signal-to-interference-plus-noise ratio (SINR)
- SIR-B. *See* Shuttle Imaging Radar B (SIR-B)
- SISO system. *See* Single-input, single-output (SISO) system
- Skin return, 543
- Slant/ground plane operations, stripmap SAR, 323–324
- Slant plane image, 236
- SLB. *See* Sidelobe blanking (SLB)
- SLC. *See* Sidelobe cancellation (SLC)
- SLE. *See* Integrated sidelobe energy (SLE)
- Slow-moving target detection, GMTI radars and, 722, 722*f*
- Slow time, 155
- Slow-time
 - defined, 720
- Slow-time space-time coding. *See* Interleaved scanning
- SMA. *See* Scattering matrix with absolute phase (SMA)
- Small-signal suppression effect, 569
- SMART-L surveillance radar, 410, 411*f*
- SMR. *See* Scattering matrix with relative phase (SMR)
- SN. *See* Spot noise (SN)
- Sniff mode (sniffer), 572
- SNR. *See* Signal-to-noise ratio (SNR)
- Soft thresholding operation, 173–175
- SOJ. *See* Stand-off jammer (SOJ)
- Solid-state amplifier (SSA), 540
- SORO. *See* Scan-on-receive-only (SORO)
- SP. *See* Subspace pursuit (SP)
- Space and Aerospace Systems (SAS), 707
- Spaceborne systems, 384–385
 - vs. airborne, 360–361
- Space-CMA (S-CMA), 769, 770, 771
- Space domain, clutter rejection in, 518–520, 518*f*, 519*f*
- Space-time adaptive processing (STAP), 1, 7, 11, 87, 155, 454, 707
 - angle estimation, 490
 - for clutter rejection, 499

- computational demands, 493
- covariance estimation, 455
- fundamentals, 478–483
- GSC. *See* Generalized sidelobe canceller (GSC)
- heterogeneous clutter, 492
- implementation issues, 491–493
- interleaved scanning, 516, 516*f*
- maximum SINR filter, 478–481
- mitigate clutter effects with, 722–724, 724*f*
- MV space-time beamformer, 481–482
- nonstationary clutter, 492
- notions and operations, 436–437
- overview, 454–455
- post-Doppler, 487–488, 487*f*
- pre-Doppler, 488–489, 488*f*
- radar detection block diagram, 490–491, 491*f*
- RD-STAP. *See* Reduced-dimension STAP (RD-STAP)
- reduced-rank, 489
- Space-time-CMA (ST-CMA), 769, 770, 771
- Space-time coding, 502–508. *See also* Colored transmission
 - for active antenna systems, 499–527
 - clutter rejection, Clutter rejection code selection, 517–520
 - coding strategy, 520, 520*t*
 - colored space-time exploration. *See* Colored space-time exploration
 - grating lobes effects. *See* Grating lobes
 - interleaved scanning (slow-time space-time coding). *See* Interleaved scanning
 - overview, 499–500
 - wideband MTI, 520–524, 521*f*, 522*f*, 523*f*
- Space-time processing, 474–478
 - detection, 476–477
 - improvement factor, 478
 - performance metrics, 476–478
 - SINR loss, 477–478
- Space-time signals, 466–472
 - covariance matrix, 469
 - 2D-Fourier transform (FFT), 468
 - MVDR spectrum, 470, 470*f*
 - power spectral density, 467, 467*f*
 - space-time snapshot, 468
 - vector, 466–467
- Space-time steering vector, 157
- Sparse MIMO array, 140
- Sparse reconstruction (SR), 8
 - algorithms, 166–183
 - Bayesian approaches, 178–180, 180*f*
 - Greedy algorithms, 176–178, 178*f*
 - iterative reweighting schemes, 175–176
 - matrix uncertainty/calibration, 182–183
 - overview, 147
 - penalized least squares problem. *See* Penalized least squares problem
 - radar applications, 183–195
 - structured sparsity, 180–182
 - thresholding algorithms, 172–175
- Sparsity, 151
 - block, 181
 - linear model regularization through, 158–159
 - structured, 180–182
- Sparsity-cognizant total least squares (STLS), 182
- Spatial sampling, 459–464
- Spatial steering vector, 462
- SPECAN, 273, 325
- Speckle, 5
- Spectrogram analysis, of human returns, 719–721, 721*f*
- Spectrums, 1, 6–8, 215
 - colored noise interference, 94*f*
- SPGL1 algorithm, 170–171
- SPJ. *See* Self-protection jammer (SPJ)
- Spoiled phased array, 129
 - signal correlation matrix of, 128
- Spotlight imaging, stripmap SAR, 324–325
- Spotlight mode, SAR, 8–9, 211, 220–225, 221*f*, 222*f*, 223–224*f*, 225*t*
 - parameters for, 225*t*
- Spot noise (SN), 540, 541
- Squint angle, 221*f*, 224
- Squinted operation, stripmap SAR, 325–327
- SR. *See* Sparse reconstruction (SR)
- SRC. *See* Secondary range compression (SRC)
- SRC, Inc., 706
- SRTM. *See* Shuttle Radar Topography Mission (SRTM)
- SSA. *See* Solid-state amplifier (SSA)
- SSJ. *See* Self-screening jammer (SSJ)
- SSM. *See* State space method (SSM)
- Stable local oscillator (STALO), 620
- STALO. *See* Stable local oscillator (STALO)
- Stand-forward jammer (SFJ), 534
- Stand-in jammer (SIJ), 534
- Stand-off jammer (SOJ), 534
- Standoff range, SAR, 232
- STAP. *See* Space-time adaptive processing (STAP)
- STAR-3i system, 385
- State space method (SSM), 612
- State vector, 669
- Statistical training libraries, SAR, 646
- ST-CMA. *See* Space-time-CMA (ST-CMA)
- Steered response, as pattern of array antenna, 132–133, 133*t*
- Steering vector, 125, 156
 - space-time, 157
- Stein's lemma, 636
- Stepped chirp waveform, 7, 20, 21, 40–48
 - FD approach for processing, 43–45, 45*f*
 - reception of, 41–42
 - TD approach for processing, 42–43
 - transmission, 40, 41*f*
 - variables, 23
- Stepped frequency (SF) waveform, 7, 21, 58–70, 59*f*
 - description, 58, 58*f*
 - disadvantages, 58
 - Doppler shift and, 68–69
 - DTFT, 62–63
 - parameters, 68
 - phase rotation, 59
 - processing architecture, 60, 60*f*
 - properties, 58
 - pulse width in target- or clutter-rich environment, 66–68, 67*f*
 - range ambiguities, 65–66, 65*f*
 - range compression, 60–62, 62*f*
 - range sidelobe suppression, 63
 - Rayleigh range resolution, 62
 - receiver, 60, 60*f*
 - straddle loss, 63–65, 64*f*, 65*f*
 - variables, 24
- STLS. *See* Sparsity-cognizant total least squares (STLS)
- Stokes parameters, in polarization, 598–601
- Stolt interpolation, 307, 313–315, 314*f*
- Straddle loss, 63–65, 64*f*, 65*f*
- Straight-through repeater, 540
- Stretch processing, 7, 21, 26–40, 446–449, 447*f*, 448*f*. *See also* Linear frequency modulation (LFM)
 - advantage of, 34
 - bandwidth processing, 26–27
 - implementation, 27–36
 - LFM, 19–20, 27–28
 - processor architecture, 30–31, 31*f*
 - pulse compression gain, 37–38
 - range-Doppler coupling, 39
 - range resolution, 32–33
 - range-sidelobe suppression, 38–39, 38*f*
 - receiver, 28–30
 - variables, 22–23
 - vs. correlation processing, 36–37, 37*t*
 - for wideband LFM pulse compression, 575
- Stripmap SAR, 8–10, 260–261
 - along-track, 265
 - antenna area and system utility, 322–323
 - applications
 - penetration applications, 329–330, 330*f*

- Stripmap SAR (*cont.*)
 remote sensing, 327–329, 328*t*, 329*f*
 cross-range, 265
 down-range, 265
 image formation survey, 269
 algorithms, 268*t*
 integration angle, 265–266, 265*f*
 point spread response (PSR), 266–267, 267*f*, 269. *See also* Point spread response (PSR)
 pulse repetition frequency (PRF)
 lower limits, 318–321
 upper limits, 321–322
 RDA. *See* Range-Doppler algorithms (RDA)
 RMA. *See* Range migration algorithm (RMA)
 RSA. *See* Range stacking algorithm (RSA)
 slant and ground planes, 323–324
 spotlight image formation techniques, 324–325
 squinted operation, 325–327
 STS-12000 Long-Range Perimeter Surveillance Radar, 706
 Subarrays, digital beamforming (DBF), 411–414
 architecture, 412–413, 413*f*
 irregular, 412–413, 413*f*
 overlapped, 413, 413*f*, 414, 417–418, 417*f*
 regular, 412, 413, 413*f*
 array factor and subarray pattern, 413–414, 413*f*–414*f*, 416*f*
 digitally resteeered antenna pattern, 415–419, 416*f*
 hybrid scheme of analog beamforming, 411, 412*f*
 layout choice, 412
 mapping, 411–412
 uniform linear array (ULA), 414–415, 415*f*
 Sub-arrays, grating lobes and, 518–520, 518*f*, 519*f*
 Subarray space adaptive canceller, 441
 Subspace pursuit (SP), 177, 178
 Superresolution, 643
 Support vector machines (SVMs)
 SAR, 651
 Surface acoustic wave (SAW) delay lines, 538
 SVMs. *See* Support vector machines (SVMs)
 Swath depth/width, 7, 224–225
 SWC. *See* Scan with compensation (SWC)
 Sweep rate, 226
 Swept AM (SAM), 544, 559
 Synthetic aperture radar (SAR), 1, 2, 4–5, 259–260, 499, 611, 707
 airborne collection geometry, 221*f*
 along-track sampling rate, 138–139
 assumptions of, 220
 autofocus techniques, 251–253
 blurring in, 29
 collection geometry, 4
 cross-range resolution, 234
 data link issues, 183
 DBS for. *See* Doppler beam sharpening (DBS), for SAR
 deramp reception, 226–229
 Fourier transform, 214–215, 216*t*
 identifying target set, 638–639, 639*f*, 639*t*
 imaging. *See* Imaging, SAR
 interferometric, 10
 MIMO, 138–140, 140*t*
 multiplicative noise ratio, 9
 observing feature set, 643–646
 image formation and pre-processing techniques, 643–644
 observation conditions, 643
 pre-screening and discrimination algorithms, 644–646
 overview, 211–212
 phase error effects, 244–250
 plane waves, 218–220, 219*f*
 PRF, 231–233, 233*f*
 PSR for. *See* Point spread response (PSR)
 sampling and resolution, 225–234
 temporal, 226
 selecting feature set, 639–642
 additional features, 641
 contrast features, 641
 feature dependency on angle, 641–642, 642*f*
 polarimetric features, 640
 size features, 641
 texture features, 641
 types, 640
 sinc function, 215–218, 217*f*
 spatial frequency, 218–220, 219*f*
 spotlight mode, 8–9, 211, 220–225, 221*f*, 222*f*, 223–224*f*, 225*t*
 stripmap mode, 8–10
 testing feature set, 646–651
 adaptive feedback, 647
 Bayesian network, 650
 correlation filter, 651
 distance classifier correlation filter (DCCF), 651
 distances between features, 648–649
 evolutionary programming, 648
 hidden Markov model (HMM), 650–651
 hybrid (image library and model), 648
 invariant histograms, 649
 likelihood-based classifier, 649
 maximum average correlation height (MACH), 651
 mean-squared error (MSE), 648
 models, 647
 nearest neighbor, 650
 neural nets, 649–650
 persistent scatterers, 647–648
 statistical training libraries, 646
 support vector machines (SVMs), 651
 template libraries, 647
 3-D imaging, 187–188, 188*f*
 Synthetic aperture sonar (SAS) system, 140
 Syracuse Research Corporation, 707
 Systematic errors, 381–382, 382*f*
 sensor geometry, 381
 System sensitivity, 552
- T**
 TanDEM-X, 385
 Target coherence, 509–511, 510*f*, 511*f*
 Target gain, 546
 Target helicity angle, 605
 Target polarizability angle, 605
 Targets
 determination, optimum MIMO waveform design for, 99–104, 100*f*, 101–103*f*
 in radar polarimetry, 611–612
 feature extraction, 612–616, 613*f*, 614*f*, 615*f*, 616*f*
 Target skip angle, 605
 Target tracking
 acronyms, 670–671
 covariances, 671, 671*f*
 key points, 669–670
 measurement-to-track data association, 672–675, 672*f*, 672*t*
 assignment algorithms, 674–675
 cost functions, 674
 gating, 673–674
 multisensor tracking, 691–694. *See also* Multisensor tracking
 multitarget tracking, 677–690. *See also* Multitarget tracking
 notation, 670
 organization, 669
 track filtering, 675–677, 676*f*
 Target transfer matrix, 89, 90, 93
 Taylor series, 50
 Taylor weighting, 52–53, 54*f*, 56
 TB. *See* Terrain bounce (TB)
 T-CMA. *See* Time-only CMA (T-CMA)
 TD. *See* Time domain (TD)
 TDM. *See* Time division multiplexing (TDM)

- TED. *See* Total energy detector (TED)
- Template libraries, SAR, 647
- Temporal decorrelation, InSAR, 378, 379*f*
- Temporal sampling, 464–466
synthetic aperture radar, 226
- Temporal steering vector, 465
- Terrain bounce (TB), 544
- Terrain map estimation, InSAR, 374
- Terrain motion mapping, 386–388, 387*f*, 388*f*
- TerraSAR-X, 385
satellite images, 221, 223*f*
- THALES, 410, 411*f*
- Thermal and quantization noise, InSAR, 376–377
- Thermal noise, 242
formulas, 545–546
- 13-element Barker code
conversion to quadriphase code, 71*t*
- 3-D SAR imaging, 187–188, 188*f*, 358–359
- 3-D scatterer models, ISAR, 654
- Threshold detector, 537
- Thresholding algorithms, 172–175
hard thresholding operation, 175
soft thresholding operation, 173–175
- Throughput delays, 564
- Tikhonov formulation, 190, 193, 194*f*
- Tilt angle (ψ), 596, 597, 598
- Time-bandwidth product, 545
- Time-delay steering, DBF, 405–406
- Time division multiplexing (TDM), 135
- Time-domain phase function, 51, 56
- Time domain (TD)
clutter rejection in, 517, 517*f*
envelope, 51
stepped chirp waveform processing, 42–43
- Time of arrival (TOA), pulse, 537, 538
- Time-only CMA (T-CMA), 769, 770, 771
- Time-varying ranges
with Boulic-Thalmann model, 716–718, 717*t*
- TOA. *See* Time of arrival (TOA)
- TOMHT. *See* Track-oriented MHT (TOMHT)
- Tomographic paradigm, SAR imaging, 234–235, 235*f*
- Tomographic SAR, 359–359
- Topologies, radar system, 5–6
- Total composite channel transfer matrix, 94
- Total energy detector (TED), 572
- Towed RF decoy (TRD), 545
- TPM. *See* Transition probability matrix (TPM)
- TPS. *See* Transport parameter signalling (TPS)
- Track deception/denial, 542, 543–544
- Track filter acceleration limit, 579–580
- Track filtering, 669, 675–677, 676*f*
extended Kalman filter, 13
- Track-level fusion, 692–693, 693*f*
challenges, 692–693, 694
- Track-on-jam, 580
- Track-oriented MHT (TOMHT), 685–686
- Transfer matrix, 98
- Transition probability matrix (TPM), 689
- Transmit-independent channel estimation, 110–111, 110*f*, 111*f*
- Transmit power reduction, 561–562
- Transmitter-based EP, 561–562
EMCON, 562
high peak power, 561
transmit power reduction, 561–562
- Transmitter peak power, high, 561
- Transmitters, PBR, 739–740
- Transponder, 536
- Transport parameter signalling (TPS), 787
- Traveling-wave tube (TWT) amplifiers, 536, 540
- T/R isolation, 537
- TV norm, 171–172, 193, 194*f*
- Two-dimensional cross-correlation function (2D-CCF), in PBR, 740–741
FFT, efficient implementations based on
correlation FFT, 749–750, 749*f*, 750*t*
direct FFT, 749*f*, 750–751, 751*t*
range sidelobe control
in WiFi-based PBR, 776–780, 777*f*, 778*t*, 780*f*
in WiMAX-based PBR, 780–786, 781*t*, 782*f*, 783*f*, 785*f*
shape control for DVB-T signals, 786–791, 787*f*–790*f*, 787*t*
suboptimum implementations, 751–755
batches algorithm, 751–753, 752*f*, 753*f*
channelization technique, 753–755
- Two-dimensional (2D) phase unwrapping, InSAR, 366–373
differences and commonality, 372–373
least squares method, 369–371, 370*f*, 371*f*
network flow method, 371–372
path-following approach, 366–369, 368*f*, 369*f*
- 2-D SAR images, 344–345, 345*f*
- U**
- UASs. *See* Unmanned aerial systems (UASs)
- UHF FOPEN SAR, 323
- ULA. *See* Uniform linear array (ULA)
- Ultra-wideband SAR, 640
- UMTS. *See* Universal Mobile Telecommunications System (UMTS)
- Undersampling ratio, 151
- Unequispaced FFT (USFFT), 186
- Uniform linear array (ULA), 96–97, 155
- Unit balls, 151, 151*f*
- Universal classifier, 645
- Universal Mobile Telecommunications System (UMTS), 740
- Unmanned aerial systems (UASs), 707
- Unwrapped phase estimation, InSAR, 374
- Upper limits, PRF, 321–322
- U.S. Advanced Research Projects Agency (ARPA), 385
- U.S. Jet Propulsion Laboratory, 385
- USFFT. *See* Unequispaced FFT (USFFT)
- V**
- VADER. *See* Vehicle and Dismount Exploitation Radar (VADER)
- Variables
mismatched filters, 25
NFLM, 23–24
quadriphase codes, 24–25
SF waveform, 24
stepped chirp waveforms, 23
stretch processing, 22–23
waveforms, 21–22
- Variable-structure interacting multiple model (VS-IMM) estimators, 675, 690
- Vector quantization, 644
- Vehicle and Dismount Exploitation Radar (VADER), 707
- Velocity bin masking, 542
- Velocity false targets (VFT), 542, 543
- Velocity gate capture, 543–544
- Velocity gate pull-in (VGPI), 543–544
- Velocity gate pull-off (VGPO), 543–544, 575
- Velocity gate stealing (VGS), 543–544
- Velocity gate walk-off (VGWO), 543–544
- Velocity synthetic aperture radar (VSAR) system, 722
- Velocity track deception technique, 543–544
- Vernier array, 128, 139–140
- VFT. *See* Velocity false targets (VFT)
- VGPI. *See* Velocity gate pull-in (VGPI)
- VGPO. *See* Velocity gate pull-off (VGPO)
- VGS. *See* Velocity gate stealing (VGS)
- VGWO. *See* Velocity gate walk-off (VGWO)
- Virtual array, MIMO radar, 122–124, 124*f*
properties of, 140*t*
- VSAR system. *See* Velocity synthetic aperture radar (VSAR) system
- VS-IMM. *See* Variable-structure interacting multiple model (VS-IMM) estimators

- W**
- Wald sequential test, comparison of algorithms via, 637
 - Walk phase, in pull-off sequence, 544
 - Waveform generators (WFG), 404–405
 - Waveforms, 1, 6–8
 - biphase-coded, 70
 - composite, 40, 42
 - LFM, 19–20
 - for MIMO radar, 135–138, 136–137*f*
 - myriad, 226
 - NLFM. *See* Nonlinear frequency modulated (NLFM) waveforms
 - orthogonal. *See* Orthogonal waveforms
 - phase-coded, 20
 - pulse compression. *See* Pulse compression (PC), waveform
 - quadruphase, 7, 70–75
 - resolution, 7
 - stepped chirp. *See* Stepped chirp waveform
 - stepped frequency. *See* Stepped frequency (SF) waveform
 - time-domain envelope, 51
 - variables, 21–22
 - Wavelet transform, 655
 - Wave polarization, 589
 - fundamentals of, 594–598, 594*f*
 - circular, 595, 595*f*
 - elliptical, 596–598, 596*f*
 - linear, 595, 595*f*
 - Weather effects, InSAR, 378, 379
 - Weight estimation, adaptive jammer cancellation, 429–432
 - Cholesky factorization, 432
 - key considerations, 429–430
 - sample matrix inversion, 430–431
 - SINR loss, 430, 431*f*
 - weight jitter stabilization, 431–432, 431*f*
 - Weighting network
 - for Barker sidelobes reduction (BWN), 777–778, 778*t*
 - Weight jitter
 - defined, 431
 - stabilization, 431–432, 431*f*
 - WFG. *See* Waveform generators (WFG)
 - Whitening matched filter, 90
 - output of, 91
 - White scatterers model, 509n6
 - Wideband limiting, 570–571
 - Wideband MTI, 520–524, 521*f*, 522*f*
 - ecartometry measurements, 523, 523*f*
 - MiTI, 523–524, 523*f*
 - Wideband operation, for adaptive beamforming, 441–449
 - antenna pattern response, 442*f*
 - jammer dispersion, 443, 444*f*
 - joint spatial-frequency domain, 444–446
 - LFM/stretch, 446–449, 447*f*, 448*f*
 - subbanding, 446, 446*f*
 - Wideband phase error, 248–250, 249*f*
 - Wideband pulse compression waveforms, 565
 - Wide instantaneous dynamic range, 569
 - Wide-pulse waveform, 562–563
 - Wiener filter, 422–423
 - WiFi-based PBR, range sidelobe control in, 776–780, 777*f*, 778*t*, 780*f*
 - WiFi transmitters, 740
 - Wiley, Carl, 211
 - WiMAX. *See* Worldwide Interoperability for Microwave Access (WiMAX)
 - WiMAX-based PBR, range sidelobe control in, 780–786, 781*t*, 782*f*, 783*f*, 785*f*
 - Worldwide Interoperability for Microwave Access (WiMAX), 740
 - Wrapped interferometric phase difference, InSAR, 365–366
 - Wrapped phase, InSAR, 351–354, 353*f*
- X**
- X band system, 385
 - XEYE. *See* Cross-eye (XEYE)
 - Xpatch Electromagnetic Simulation Software, 638
 - XPOL. *See* Cross-polarization (XPOL)
- Z**
- Zebra plots, 233
 - Zero padding, 44, 63

Principles of Modern Radar: Advanced Techniques

This second of three volumes in the **Principles of Modern Radar** series offers a much-needed professional reference for practicing radar engineers. It provides the stepping stones under one cover to advanced practice with overview discussions of the most commonly used techniques for radar design, thereby bridging readers to single-topic advanced books, papers, and presentations.

It spans a gamut of exciting radar capabilities from exotic waveforms to ultra-high resolution 2D and 3D imaging methods, complex adaptive interference cancellation, multi-target tracking in dense scenarios, multiple-input, multiple-output (MIMO) and much more.

All of this material is presented with the same careful balance of quantitative rigor and qualitative insight of *Principles of Modern Radar: Basic Principles*. Each chapter is likewise authored by recognized subject experts, with the rigorous editing for consistency and suggestions of numerous volunteer reviewers from the radar community applied throughout.

Advanced academic and training courses will appreciate the sets of chapter-end problems for students, as well as worked solutions for instructors. Extensive reference lists show the way for further study. It is an excellent choice for a second course in radar after mastering *Principles of Modern Radar: Basic Principles* (SciTech, 2010).

KEY FEATURES

- Each chapter written and illustrated by recognized, authoritative experts on the topic
- Problems and extensive reference lists provided to test understanding and encourage further study
- Reviewed by experts within the radar community and rigorously edited for coherence and consistency

OTHER VOLUMES IN THE PRINCIPLES OF MODERN RADAR SERIES

Principles of Modern Radar: Basic Principles

- edited by Mark A. Richards, James A. Scheer, and William A. Holm • ©2010 • ISBN 978-1891121-52-4

Principles of Modern Radar: Radar Applications

- edited by James A. Scheer and William L. Melvin • ©2013 • ISBN 978-1-891121-54-8

ABOUT THE EDITORS

Dr. William L. Melvin is Director of the Sensors and Electromagnetic Applications Laboratory (SEAL) at the Georgia Tech Research Institute and an Adjunct Professor in Georgia Tech's Electrical and Computer Engineering Department. He has authored over 150 publications and holds three patents on adaptive radar processing.

James A. Scheer has been directly involved in radar research and development for over 40 years. He is editor of the book *Coherent Radar Performance Estimation* (Artech, 1993), has authored chapters in five radar textbooks, and is an instructor in a variety of radar short courses, including Principles of Modern Radar.



Edison, New Jersey
www.scitechpub.com

